THE DYNAMICS OF GEODESIC FLOWS*

GUSTAV A. HEDLUND

1. **Introduction.** Geodesic systems, particularly those on two-dimensional manifolds, have been a rich source in the determination and display of the possible types of macroscopic behavior of the motions of dynamical systems. In connection with the question of the existence of periodic motions, Poincaré [1]† investigated the geodesics on convex surfaces. Hadamard [1] has constructed open surfaces of negative curvature and proved the existence of interesting classes of geodesics on these surfaces. By an ingenious use of symbolism to characterize these geodesics, Morse [1] proved the existence of nonperiodic recurrent geodesics of discontinuous type. Birkhoff [1] has constructed closed surfaces of nonpositive curvature and has shown that, among many other types, there exist transitive geodesics on these surfaces.

There is another group of mathematicians who have made numerous contributions in connection with geodesic systems on surfaces of constant negative curvature. As will be seen, these surfaces have a close relationship with Fuchsian groups, and in addition to their work having other connections with these groups, Artin [1], Myrberg [1, 2, 3], Nielsen [1, 2], Koebe [1], and Löbell [1, 2, 3, 4] have derived many properties of the geodesics.

With the recent developments in ergodic theory, interest has been centered on those properties of geodesic flows associated with transitivity in some form, as for example, regional transitivity, metric transitivity, and mixture. The conditions under which regional transitivity holds have been greatly extended by Morse [3]. Geodesic systems have furnished some of the few known examples of metrically transitive dynamical systems (cf. Hedlund [1, 2], E. Hopf [1]). As will be indicated, a number of new results concerning transitivity can be added, both in the case of constant curvature and in the case of variable curvature.

An enormous body of results has been attained, and an hour is entirely inadequate to permit a description of all. For this reason I propose to restrict the discussion to transitivity properties of geodesic flows. It has been conjectured (Birkhoff [3], p. 370) that these are the important properties in that they are *general* in some

* An address delivered before the New York meeting of the Society on October 27, 1938, by invitation of the Program Committee.

† Numbers in brackets refer to the bibliography at the end of the paper.

sense, but the evidence gathered to date is not at all conclusive. Outstanding problems remain unsolved, a notable one being the problem of metric transitivity of the geodesic flow on a closed analytic surface of variable negative curvature.

It is of interest to note the influence of Poincaré in many phases of our considerations. It was Poincaré who first emphasized macro-analysis and probability considerations in the study of dynamical systems. We are indebted to him for the Poincaré fundamental group and much of our knowledge of Fuchsian groups which play an important role in the analysis of geodesic flows on two-dimensional manifolds of negative curvature.

2. **Flows and transitivity types.** The geodesics on a Riemannian manifold are the solutions of the Euler equations, a system of second order differential equations derived by imposing the condition that the first variation of the arc length vanish. If the coefficients of the positive definite quadratic forms

$$g_{\alpha\beta}du^{\alpha}du^{\beta}$$

which define the Riemannian manifold locally are of class C'' (somewhat less than this is sufficient), a geodesic is uniquely determined by an element, that is, by a point and a direction at that point. Let us assume that the manifold is complete in the sense that each geodesic can be continued to infinite length (cf. H. Hopf and Rinow [1]). Then if g is the directed geodesic determined by the element e, and s is the sensed arc length on g measured from the point P at which e is situated, corresponding to s there is an element e_s, namely, the element of g at distance s along g from P. Furthermore, e_s varies continuously with e and s. The transformation $e \to e_s$ is a transformation T_s of the space Ω of elements on the manifold into itself, and the properties which we consider can be most simply stated in terms of such a one-parameter set of transformations T_s.

The conditions which will be imposed on the space in which the transformations are defined and on the transformations themselves will be fulfilled by the element spaces and transformations in them which we subsequently consider.

Let Ω be a metric, separable, complete space in which an l-measure in the sense of von Neumann (cf. von Neumann [1], p. 575) is defined. The measure defined by use of this l-measure, which is analogous to a Lebesgue outer measure, will be denoted by m. Let T_t be, for each real t, a one-to-one transformation of Ω into itself satisfying the following conditions:

(a) $T_0(P)=P$; $T_t[T_s(P)]=T_{t+s}(P)$.

(b) $T_t(P)$ is a continuous function of t and P.

(c) If A is a measurable subset of Ω, then $T_t(A)$ is measurable and $m[T_t(A)]=mA$.

Such a continuous one-parameter group of transformations will be called a *flow* in Ω.

It will be convenient to denote the set $T_t(A)$, $(A \subset \Omega)$, simply by A_t. The set A will be said to be *invariant* if A_t coincides with A for all t. The set P_t, $(-\infty<t<+\infty)$, will be called a *motion* or *trajectory*.

The properties of the flow T_t in Ω with which we shall be concerned are the following, where the set of points common to the sets A and B of Ω is denoted by $A\cdot B$ and the empty set is denoted by 0.

REGIONAL TRANSITIVITY. *Given D and D^*, arbitrary open sets in Ω, there exists a t such that $D_t\cdot D^*\neq 0$.*

TOPOMETRIC TRANSITIVITY. *Given M, any measurable set of positive measure in Ω, and D, any open set in Ω, there exists a t such that $M_t\cdot D\neq 0$.*

METRIC TRANSITIVITY. *Given M and M^*, arbitrary measurable sets of positive measure in Ω, there exists a t such that $M_t\cdot M^*\neq 0$.*

PERMANENT REGIONAL TRANSITIVITY. *Given D and D^*, arbitrary open sets in Ω, there exists a $\bar{t}$ such that $D_t\cdot D^*\neq 0$, $(|t|\geqq\bar{t})$.*

PERMANENT TOPOMETRIC TRANSITIVITY. *Given M, any measurable set of positive measure in Ω, and D, any open set in Ω, there exists a $\bar{t}$ such that $M_t\cdot D\neq 0$, $(|t|\geqq\bar{t})$.*

PERMANENT METRIC TRANSITIVITY. *Given M and M^*, arbitrary measurable sets of positive measure in Ω, there exists a $\bar{t}$ such that $M_t\cdot M^*\neq 0$, $(|t|\geqq\bar{t})$.*

MIXTURE. *Given M, M^*, and $\overline{M}$, arbitrary measurable sets of Ω of finite positive measure,*

$$\lim_{t\to\pm\infty}\frac{m(M_t\cdot M^*)}{m(M_t\cdot\overline{M})}=\frac{mM^*}{m\overline{M}}.$$

Regional transitivity is sometimes given a different but equivalent definition. A motion will be called *transitive* if the points on it form a set which is everywhere dense in Ω. *Regional transitivity is equivalent to the property that there exist a transitive motion* (cf., for example, Birkhoff [1], chap. 7).

The points on the transitive motions in Ω form a set which is the product of a denumerable set of open sets and is thus measurable. If this set coincides with Ω except for a set of measure zero, it will be said that *almost all motions are transitive. Topometric transitivity is a necessary and sufficient condition that almost all motions be transitive.*

Metric transitivity is a necessary and sufficient condition that in any division of Ω into two complementary, invariant, measurable sets, one of the sets is of measure zero. In this form the notion of metric transitivity was introduced by Birkhoff and Smith (cf. Birkhoff [4], p. 365). It plays a fundamental role in connection with ergodic theory. According to the ergodic theorem of Birkhoff [2], if $m\Omega$ is finite and M is any measurable set in Ω, the mean time of sojourn of a motion in M (that is, $\lim_{\beta-\alpha\to\infty} L_{\alpha,\beta}(P, M)/(\beta-\alpha)$, where $L_{\alpha,\beta}(P, M)$ is the linear measure of the part of the set P_t, $(\alpha\leqq t\leqq\beta)$, in M) exists except for a set of motions of measure zero. If metric transitivity holds, the mean time of sojourn in M is the same for almost all motions and is equal to $mM/m\Omega$.

If $m\Omega$ is finite, by replacing $\overline{M}$ by Ω the mixture property becomes

$$\lim_{t\to\pm\infty} m(M_t\cdot M^*) = \frac{mMmM^*}{m\Omega},$$

and, conversely, this implies the mixture property. Thus, any measurable set M of positive measure tends, with increasing or decreasing time, to occupy a definite fractional part of any other measurable set M^*, and the fraction is simply the fractional part of Ω which M occupies. Sets tend towards *homogeneous distribution* in Ω. (In this connection cf. E. Hopf [2], where references to the work of Koopman and von Neumann will be found.)

There are a number of evident relationships between the transitivity properties which have been defined. Metric transitivity implies topometric transitivity, which, in turn, implies regional transitivity. Any one of the permanent types of transitivity implies the corresponding non-permanent type. Permanent metric transitivity implies permanent topometric transitivity, which, in turn, implies regional transitivity. Mixture implies permanent metric transitivity and thus implies all the types of transitivity which have been defined here.

The flow defined by a suitably chosen family of parallel straight lines on a torus (rectangle with opposite sides identified), shown to be metrically transitive by Birkhoff and Smith (cf. Birkhoff [4], p. 368) yields an example which is not permanently regionally transitive and thus has none of the permanent transitivity properties. It seems to be difficult to give examples of flows which have one of the

non-permanent types of transitivity without having the other non-permanent types, or which have one of the permanent types of transitivity without possessing the other permanent types. However, it will be possible to give an example of a geodesic flow which is regionally transitive (and even permanently regionally transitive) but not metrically transitive.

3. **Two-dimensional manifolds of constant negative curvature.** The simplest manifolds on which the geodesics display transitivity properties of the kind we are considering are two-dimensional manifolds of constant negative curvature. To define such manifolds, let Ψ be the interior of the unit circle U: $x^2+y^2=1$. To Ψ we assign the metric

$$ds^2 = \frac{4(dx^2 + dy^2)}{c(1 - x^2 - y^2)^2} = \frac{4\,|\,dz\,|}{c(1 - z\bar{z})^2}, \qquad c > 0. \tag{3.1}$$

The curvature of this simply connected Riemannian manifold is $-c$. The metric (3.1) assigns a length to curves in Ψ, and this length will be called *hyperbolic length*. Angle is euclidean angle, and the element of (hyperbolic) area is

$$\frac{4dxdy}{c(1 - x^2 - y^2)^2}. \tag{3.2}$$

The geodesics defined by (3.1) are arcs of circles orthogonal to U and will be called *hyperbolic lines*. Given two points P and Q of Ψ, there is a unique hyperbolic line segment joining P and Q, and the hyperbolic length of this hyperbolic line segment is the *hyperbolic distance* between P and Q.

The metric (3.1) is invariant under linear fractional transformarions which take Ψ into Ψ, so that under such transformations, hyperbolic distance, angle, and area are invariant. Such a transformation is either an elliptic transformation with fixed points inverse with respect to U, a parabolic transformation with fixed point on U, or a hyperbolic transformation with fixed points on U. These transformations are rigid motions of the well known hyperbolic geometry under consideration.

Now let F be a Fuchsian group with U as principal circle. That is, F is a group of linear fractional transformations, each of which transforms U into U and Ψ into Ψ, such that F is properly discontinuous in Ψ (cf. Ford [1], p. 35, chap. 3). Two sets of points in Ψ are *congruent* if there is a transformation of F taking one of these sets into the other. Either set will be said to be a *copy* of the other.

It can be shown that corresponding to any such group F there exists a normal fundamental region R (cf. Ford [1], pp. 44, 69–70). This is a simply connected region bounded by arcs of hyperbolic lines which are congruent in pairs, such that no two interior points of R are congruent and any point of Ψ is congruent to some point within or on the boundary of R. If suitable conventions are made as to the inclusion of boundary points of R, no two copies of R have a common point and the totality of these copies fills Ψ.

If points which are congruent under F are considered identical, there is defined a two-dimensional manifold M_F^{-c} of constant negative curvature $-c$. These manifolds are non-euclidean space forms of hyperbolic type, and an extensive analysis of them can be found in the papers of Koebe [1] and Löbell [1]. By including in the group F transformations of the form

$$w = \frac{a\bar{z} + \bar{c}}{c\bar{z} + \bar{a}}, \qquad a\bar{a} - c\bar{c} = 1,$$

these authors consider non-orientable as well as orientable manifolds. For simplicity, the discussion will be restricted to manifolds defined by Fuchsian groups, though the results derived apply to the non-orientable cases.

The presence of elliptic transformations in F, such a transformation necessarily having one of its fixed points in Ψ, implies the existence of singular points on the manifold M_F^{-c}. The total angle at such a point is not 2π.

In the nonsingular case there are restrictions on the topological invariants of the manifolds. If the manifold is closed and orientable, its genus must be greater than one. (In the non-orientable case, the genus of a closed manifold must be greater than two.) Among the open manifolds are included manifolds of finite or infinite connectivity.

As a first classification of the manifolds M_F^{-c}, they are divided into first and second kind. The group F is of the first kind if it is not properly discontinuous on U. The corresponding manifold M_F^{-c} will be said to be of the *first kind* and denoted by M_I^{-c}. If F is not of the first kind, it is of the second kind, and the corresponding manifolds of the *second kind* will be denoted by M_{II}^{-c}. An essential difference between Fuchsian groups of the first and second kind lies in the behavior of the fundamental region. In the case of a group of the second kind, the fundamental region abuts on the circle U in an interval (cf. Ford [1], pp. 74–75). Thus manifolds of the second kind are necessarily open. In the case of groups of the first kind the boundary of

the fundamental region cannot contain an interval of U and may or may not have points on U. Manifolds of the first kind include all closed orientable two-dimensional Riemannian manifolds of constant negative curvature. In addition there are included manifolds which are not closed and which may be of finite or infinite connectivity.

An element e in Ψ is a point of Ψ together with a direction at that point and can be specified by three coordinates (x, y, ϕ), where x and y are the coordinates of the point and ϕ, $(0 \leqq \phi < 2\pi)$, is an angular coordinate at the point measured positively in the counterclockwise sense from a direction parallel to the positive x-axis. The point $P(x, y)$ is the point *bearing* the element (x, y, ϕ). A *neighborhood* of the element $e_1(x_1, y_1, \phi_1)$ is the set (x, y, ϕ) such that

$$H(P, P_1) < \delta, \qquad \|\phi - \phi_1\| < \delta,$$

where P is the point (x, y), P_1 is the point (x_1, y_1), $H(P, P_1)$ denotes the hyperbolic distance between P and P_1 when $c = 1$, $\|\phi - \phi_1\|$ denotes the least value of the set $|\phi - \phi_1 + 2n\pi|$, $(n = 0, \pm 1, \pm 2, \cdots)$, and $\delta > 0$. Let E denote the space of elements in Ψ with neighborhoods thus defined.

A transformation of F carries an element into a *congruent* element. The space Ω_F of elements on $M_{\bar{F}^c}$ is the space obtained by identifying congruent elements of E. Neighborhoods are defined in Ω_F as the correspondents of the neighborhoods in E, and Ω_F is a Hausdorff space. By suitably defining a metric in Ω_F, Ω_F is made a metric, separable, complete space.* The points of Ω_F are in 1-1 correspondence with a subset E_R of E obtained by restricting the points bearing the elements to the fundamental region R (a properly chosen subset of the elements at boundary points of R included). Measurability is defined in E_R by considering this as a subset of the three-dimensional euclidean space (x, y, ϕ), and measure in E_R is defined by the integral

$$\iiint \frac{4dxdyd\phi}{c(1 - x^2 - y^2)^2}.$$

Measurability and measure in Ω_F are defined by the correspondence with E_R.

The geodesics on $M_{\bar{F}^c}$ are *represented* in Ψ by sets of hyperbolic lines, congruent hyperbolic lines representing the same geodesic. The geodesics on $M_{\bar{F}^c}$ define a flow in Ω_F which can be described simply as follows. Let p be any point of Ω_F, and let e be one of the

* The space Ω_F is metrizable. It may be necessary however to modify the metric thus obtained in order to assure completeness.

congruent elements in Ψ determining p. The element e determines a directed hyperbolic line h. Let s be the sensed hyperbolic length on h measured from the point Q bearing e. Let e_s be the element of h at the point with coordinate s, and let p_s be the point of Ω_F determined by e_s. The transformation $p \to p_s$ is a 1-1 continuous, measure preserving transformation T_s of Ω_F into itself. The flow thus defined is the flow in Ω_F which we consider and will be called the *geodesic flow* on M_F^{-e}.

A fundamental result and one which is useful in the derivation of transitivity properties is the following:

THEOREM 3.1. *There exist denumerably many periodic geodesics on any M_Γ^{-e}, and the elements on these geodesics form a set which is everywhere dense in Ω_F.*

A periodic geodesic on M_Γ^{-e} is represented in Ψ by an axis (the hyperbolic line joining the fixed points) of a hyperbolic transformation of F, and the statement of the theorem is equivalent to the statement that given arbitrary intervals I_1 and I_2 of U, there exists a hyperbolic transformation of F with one fixed point in I_1 and with the other fixed point in I_2.

This result was first proved in various special cases by Artin [1] and Herglotz, J. Nielsen [1], and Morse [2]. The general result is due to Koebe [1] and Löbell [2]. The proof of it is attained by simple geometrical arguments.

With the aid of the preceding theorem it is easily shown that there exists a transitive geodesic, and the following theorem can be stated:

THEOREM 3.2. *The geodesic flow on any M_Γ^{-e} is regionally transitive.*

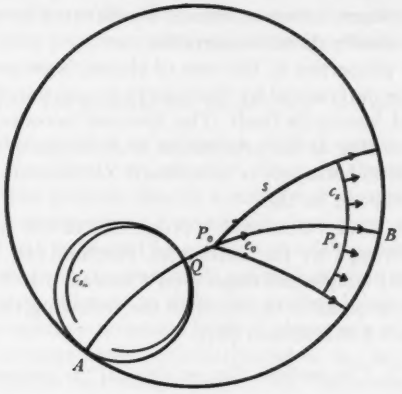
As in the case of the preceding theorem, this theorem was proved in various special cases by Artin [1] and Herglotz, by Myrberg [1, 2], and by J. Nielsen [1, 2], while the proof in the general case is due to Koebe [1] and Löbell [2].

The problem of permanent regional transitivity on any M_Γ^{-e} can be solved by an analysis of the transitivity properties of the horocycles. Since the method of solving this problem is characteristic of the methods of solving a number of other transitivity problems, it will be described briefly.

A *horocycle* is an ordinary euclidean circle which is internally tangent to U. Its importance here lies in the fact that it is a limiting curve of hyperbolic circles. A *hyperbolic circle* is the locus of points at constant hyperbolic distance from a fixed point (the center) in Ψ and is also a euclidean circle. If a sequence of hyperbolic circles which

pass through a fixed point Q of Ψ and whose centers approach a point A of U is given, the limiting curve of this sequence is the horocycle which passes through Q and is internally tangent to U at A. This horocycle will be denoted by $C(Q, A)$, the point A being the *point at infinity* of $C(Q, A)$.

Permanent regional transitivity means that under the geodesic flow any open set D in Ω_F eventually (interpreting s as time) intersects any other open set D^* and after a sufficiently great length of time, positive or negative, the intersection is never empty. Assuming that F is of the first kind, let Ω_F be denoted by Ω_1. Since, according to Theorem 3.1, the elements on periodic geodesics are everywhere dense in Ω_1, D contains such a periodic element. This element is repre-



sented in Ψ by an element e_0 of a directed axis AB of a hyperbolic transformation T of the group F. Since D is an open set, it contains the set N_0 of Ω_1 determined by a sector of elements E_0 at the point P_0 bearing e_0 and with e_0 as central element. The set D_s, $(s>0)$, then contains the set N_s of Ω_1 determined by the elements E_s perpendicular to an arc c_s of a hyperbolic circle with center P_0 and hyperbolic radius s. The arc c_s is that determined by the hyperbolic rays with initial elements in E_0, and the elements E_s are directed outward.

Let Q be any point of the axis AB. The points $T^n(Q)$, $(n=0, \pm 1, \cdots)$, have A and B as limit points and the sequence $0<s_1<s_2<\cdots$, $\lim_{n\to\infty} s_n=+\infty$, can be so chosen that the point P_{s_n} where c_{s_n} cuts across AB is congruent to Q under a power T^{m_n} of T. As a matter of fact, this sequence can be taken in the form

$s_1+n\omega$, where ω is the hyperbolic distance through which T moves a point of the axis AB. The transformed arc $c'_{s_n}=T^{m_n}(c_{s_n})$ passes through Q, and as n becomes infinite, c'_{s_n} approaches the horocycle $C(Q, A)$. The transformed sets $E'_{s_n}=T^{m_n}(E_{s_n})$ have as limit elements the elements perpendicular to $C(Q, A)$ and directed outward. Congruent elements in Ψ determine the same points in Ω_I, so the set E'_{s_n} determines points in N_{s_n}, which is a subset of D_{s_n}. If the set of elements outwardly perpendicular to the horocycle $C(Q, A)$ determines an everywhere dense set in Ω_I, the set N_{s_n} will tend with increasing n towards a dense distribution in Ω_I and for sufficiently large n will contain points in D^*. If these properties hold for any choice of Q on the axis AB, it can be shown that the flow is permanently regionally transitive in the positive sense. Similar reasoning can be applied to the case $s<0$, where, however, outwardly directed normals must be replaced by inwardly directed normals.

The density properties of the sets of elements perpendicular to a horocycle can be determined by the density properties of the elements on the directed horocycle itself. The directed horocycles are called *right* or *left* according as the orientation is clockwise or counterclockwise. A right (left) horocycle is *transitive* if its elements determine an everywhere dense set in Ω_F.

The analysis of the transitivity properties of the horocycles has been carried through by the author (cf. Hedlund [4]). The methods are geometrically simple and depend on Theorem 3.1. With the aid of these results it is possible to establish the following theorem (essentially Theorem 3.1 of Hedlund [4]):

THEOREM 3.3. *The geodesic flow on any $M_{\bar{I}^c}$ is permanently regionally transitive.*

It is of interest to note that if the fundamental region lies, together with its boundary, interior to U, all the right (left) horocycles are transitive and they define a flow in the three-dimensional space Ω_I in which all the motions are transitive.

As to metrical results with respect to manifolds $M_{\bar{F}^c}$, with an exception to be noted these have been derived only under the hypothesis that F has a finite set of generators. The corresponding manifolds will be denoted by $M_{1\bar{f}^c}$. In the case of an $M_{1\bar{f}^c}$, the (hyperbolic) area of the fundamental region is finite, which implies that the corresponding element space Ω_{1f} is of finite measure. This fact seems to play an important role in the derivation of the results to be stated.

Topometric transitivity was first proved by Artin [1] and Herglotz, and by Myrberg [1] for a special group. Later Myrberg [2, 3]

showed that topometric transitivity held for a large class of manifolds $M_{1\bar{f}}{}^{e}$. But these results are all included in the following theorem which is due to E. Hopf:

THEOREM 3.4. *The geodesic flow on any $M_{1\bar{f}}{}^{e}$ is metrically transitive.*

This important result was attained by ingenious methods involving harmonic functions, and Professor Hopf described this work in an address to this society in February, 1936 (cf. E. Hopf [1]). Previous to this the author had succeeded in establishing metric transitivity for a certain denumerable subclass of the manifolds $M_{1\bar{f}}{}^{e}$. The methods used involved symbolic characterizations of the geodesics due in one case to Artin [1] and Herglotz and in the remaining cases to J. Nielsen [3].

It is now possible to add a theorem which completes the solution of our transitivity problems with respect to manifolds $M_{1\bar{f}}{}^{e}$ (cf. Hedlund [7]).

THEOREM 3.5. *The geodesic flow on any $M_{1\bar{f}}{}^{e}$ is a mixture.*

The method of proof is an extension of that used to prove permanent regional transitivity on any $M_{\bar{1}}{}^{e}$. Since, in the case of an $M_{1\bar{f}}{}^{e}$, $m\Omega_{1f}$ is finite, the geodesic flow is a mixture if measurable sets tend with increasing or decreasing time towards homogeneous distribution in Ω_{1f} (cf. §2). As indicated in connection with permanent regional transitivity, sets in Ω_{1f} tend towards a distribution in sets determined by elements perpendicular to right horocycles. It would seem likely then that the mixture property holds if these sets of perpendicular elements are in some sense equidistributed in Ω_{1f}, or, if the same is true of the elements on the right horocycles. The elements on a right horocycle determine a path in Ω_{1f}. How is it possible to determine anything about the distribution of such paths in Ω_{1f}? An obvious way is to show that these paths are the motions of a measure preserving flow in Ω_{1f} which is metrically transitive.

Similar to the way in which the directed hyperbolic lines define the geodesic flow T_t, the right horocycles define a flow ${}_RH_t$ which bears a simple relationship to the flow T_t. Due to this relationship it is possible to show that the flow ${}_RH_t$ is metrically transitive if the same is true of the geodesic flow. But, from Theorem 3.4, metrical transitivity of the geodesic flow holds on manifolds $M_{1\bar{f}}{}^{e}$; thus on such manifolds the right horocycle flow is metrically transitive. It follows from the ergodic theorem of Birkhoff that almost all the right horocycles determine paths which are equidistributed in the sense that the mean time of sojourn of a path in a measurable set $N \subset \Omega_{1f}$ exists

and is equal to $mN/m\Omega_{I}$, except for a set of paths of measure zero. With the aid of this, Theorem 3.5 can be proved directly.

An example due to Seidel [1] shows that in the case of manifolds $M_{\bar{I}}{}^{e}$, permanent regional transitivity is not necessarily accompanied by metric transitivity when the number of generators of F is not finite. Seidel gives an example of a Fuchsian group F which is regionally transitive on U but not metrically transitive on U. It follows that F must be of the first kind and the geodesic flow on the manifold $M_{\bar{I}}{}^{e}$ defined by F is permanently regionally transitive (Theorem 3.3). This flow cannot be metrically transitive, for this would imply the metric transitivity of F on U. The following theorem can be stated:

THEOREM 3.6. *There exist manifolds $M_{\bar{I}}{}^{e}$ on which the geodesic flow is permanently regionally transitive but not metrically transitive.*

As to manifolds $M_{\mathrm{II}}{}^{e}$ the fundamental region has an interval of U on its boundary and $m\Omega_{\mathrm{II}}$ is therefore infinite. It is easily shown that regional transitivity cannot hold. It can be shown that on these manifolds almost all the geodesics are unstable in the sense that for both increasing and decreasing s, they eventually leave and remain outside of any finite domain of the manifold. This was first proved for manifolds $M_{I}{}^{e}$ by E. Hopf [1]. The general result is a corollary of Theorem 4.6 of the following section.

THEOREM 3.7. *Almost all the geodesics on any $M_{\mathrm{II}}{}^{e}$ are unstable.*

4. Two-dimensional manifolds of variable negative curvature. A number of the transitivity properties of the geodesic flows on manifolds of constant negative curvature can be shown to hold on manifolds which are not of constant curvature. The most general case in which transitivity properties have been derived is under some instability condition such as that of Morse (cf. Morse [3], Hedlund [3]).

To define the manifolds which we consider we again start with the unit circle U, but now assign to its interior the metric

$$ds^2 = \frac{\lambda^2(x, y)(dx^2 + dy^2)}{(1 - x^2 - y^2)^2}, \tag{4.1}$$

where $\lambda(x, y)$ is a function of class C^7 in Ψ and is such that $0 < a \leqq \lambda(x, y) \leqq b$. The geodesics to be considered are those defined by (4.1).

Now let us assume that $\lambda(x, y)$ is invariant under a Fuchsian group F with principal circle U. Then the metric (4.1) is invariant under the

transformations of the group, and if points in Ψ which are congruent under F are considered identical, there is defined a two-dimensional Riemannian manifold $M_F(\lambda)$. These manifolds form an extensive class. They include, in particular, all closed orientable surfaces of genus greater than one and defined by functions of at least class C^3.

The space of elements on $M_F(\lambda)$ can be taken as the space Ω_F defined in the case of constant negative curvature. The definition of measure in Ω_F is like that of the case of constant curvature except that the volume element now used is

$$\frac{\lambda^2(x, y)dxdyd\phi}{(1 - x^2 - y^2)^2}.$$

The flow to be considered in Ω_F is that defined by the geodesic determined by (4.1), and it is a continuous measure preserving flow.

In the terminology of Morse [3], two curves in Ψ are of the same *type* if there exists a constant d such that any point of either one of these curves is at a hyperbolic distance less than d from some point of the other curve. It can be shown that given an arbitrary hyperbolic line h in Ψ, there exists a geodesic g defined by (4.1) such that g and h are of the same type. The geodesics defined by (4.1) in Ψ satisfy the condition of *unicity* if there is just one of the type of a given hyperbolic line. The condition of unicity is satisfied if the curvature is negative, but, more generally, as shown by Morse [3], unicity is implied by a condition of uniform instability which is defined in terms of the equations of variation of the geodesics.

Let M_F^u denote a manifold $M_F(\lambda)$ for which the condition of unicity is satisfied. A one-to-one correspondence between the geodesics on $M_{\bar{F}^c}$ and those on M_F^u can be defined by means of the correspondence between the hyperbolic lines and geodesics of (4.1) in Ψ. This correspondence preserves many properties of the geodesics. To a periodic geodesic on $M_{\bar{F}^c}$ corresponds a periodic geodesic on M_F^u; to a transitive geodesic on $M_{\bar{F}^c}$ corresponds a transitive geodesic on M_F^u. Thus if M_F^u is denoted by M_1^u when F is of the first kind, the following theorem is implied by Theorem 3.2:

THEOREM 4.1. *The geodesic flow on any M_1^u is regionally transitive.*

This theorem is essentially due to Morse [3]. It includes a result due to Birkhoff ([1], pp. 238–248), who had previously shown the existence of transitive geodesics on certain surfaces of nonpositive variable curvature. It is possible to establish regional transitivity un-

der a condition of ray instability which differs from the instability condition of Morse (cf. Hedlund [3]).

Until recently, the preceding theorem was the only transitivity property known to hold without the restriction that the curvature be constant. It is now possible to add a number of results. These have been obtained only under the assumption that the curvature of the manifold lies between two negative constants, but it seems likely that they can be extended to hold under a condition approximating that of uniform instability. The manifolds will be denoted by M_{I}^{*}, M_{II}^{*} or M_{If}^{*} according to the properties of the Fuchsian group under which (4.1) is invariant. The following theorem was proved by Grant [1]:

THEOREM 4.2. *The geodesic flow on any manifold M_{I}^{*} is permanently regionally transitive.*

This was derived by extending the notion of horocycles to the case under consideration. The hyperbolic circles are replaced by geodesic circles. If we consider a sequence of geodesic circles, all passing through a fixed point P of Ψ and with centers approaching a point A of U, the geodesic circles approach a limiting curve which can be shown to have many of the properties of the horocycles. These *generalized horocycles* will be referred to simply as horocycles.

By arguments similar to those given in the case of constant negative curvature, sectors of elements tend, under the geodesic flow on M_{I}^{*}, towards a distribution along elements perpendicular to horocycles. An analysis of the transitivity properties of the horocycles then yields the stated theorem.

These methods yield more than this, however, in the case of manifolds M_{If}^{*}. The following metrical result can be stated (cf. Hedlund [6]):

THEOREM 4.3. *The geodesic flow on any manifold M_{If}^{*} is topometrically transitive.*

Topometric transitivity is the property that any set N of positive measure in Ω_{If} eventually intersects any given open set D in a non-empty set. Such a set N does not necessarily contain all the points of Ω_{If} determined by a sector of elements in Ψ, but this condition can be approximated. Some one of the sets of elements E_{P_0} at a point P_0 of Ψ must contain a linearly measurable subset $E_{P_0}(N)$ determining points of N and of positive linear measure (that is, linear measure in terms of the angular coordinate ϕ at P_0). There must be an element e_0 belonging to $E_{P_0}(N)$ at which the linear metric density of the set $E_{P_0}(N)$ is unity. A sector with e_0 as central element can be so chosen

that the ratio of the linear measure of the subset of $E_{P_0}(N)$ in the sector to the measure of the whole sector is nearly 1.

Furthermore, since $m\Omega_{If}$ is finite, almost all points of Ω_{If} are on motions which are stable in the sense of Poisson for both positive and negative time (cf. Birkhoff [1], p. 190). This follows essentially from a well known recurrence theorem of Poincaré. Thus the element e_0 can be chosen as an element of a motion which is stable in the sense of Poisson. Let AB be the directed geodesic in Ψ determined by e_0. Let s be the sensed arc length (as measured by 4.1) on AB measured from P_0, and let e_s be the element of AB at the point with coordinate s. The Poisson stability implies the existence of a sequence $s_1 < s_2 < \cdots$, $\lim_{n\to\infty} s_n = +\infty$, such that $e_{s_n} = e_n$ has a congruent element e_n' with $\lim_{n\to\infty} e_n' = e_0$.

The arguments are now somewhat similar to those used in proving permanent regional transitivity on manifolds of constant negative curvature. Here the arc c_s is an arc of a geodesic circle, and the set N_{s_n} does not contain all the points determined by the elements outwardly perpendicular to c_{s_n}, but only those determined by elements $E_{P_0}(N)_{s_n}$ on geodesic rays with initial elements in $E_{P_0}(N)$. However, if T_n denotes the transformation of F such that $T_n(e_n) = e_n'$, it can be shown that the elements $T_n[E_{P_0}(N)_{s_n}]$ have as limit elements all the elements perpendicular to the (generalized) horocycle which passes through P_0 and has A as point at infinity. It can be shown that these perpendicular elements determine a dense set in Ω_{If}, and the stated theorem follows.

There is a difficulty in the present case which is not encountered in the case of constant curvature. This lies in the relationship between the sets of elements $E_{P_0}(N)$ and $E_{P_0}(N)_{s_n}$; more exactly, in the relationship between the linear measure of the set $E_{P_0}(N)$ and the measure (linear on c_{s_n} measured in terms of hyperbolic length) of the set bearing $E_{P_0}(N)_{s_n}$. The proof of the desired relationship is not simple and involves an analysis of the dependence of the function $G(r, \theta)$ on θ, where $G(r, \theta)$ is defined by the quadratic form

$$ds^2 = dr^2 + G^2(r, \theta)d\theta^2$$

obtained by setting up geodesic polar coordinates with P_0 as center.

The following theorem is essentially a corollary of the preceding theorem, the statement with regard to non-orientable surfaces being derivable by the same methods.

THEOREM 4.4. *Almost all the geodesics on any closed surface of class C^3 and of negative curvature are transitive.*

We state without proof the following theorem which can be derived by these methods:

THEOREM 4.5. *The geodesic flow on any closed surface of class C^3 and of negative curvature is permanently regionally transitive.*

These theorems complete the known results concerning transitivity properties of the geodesic flows on two-dimensional manifolds. Many problems remain unsolved. The manifolds M_I^n, on which, according to Theorem 4.1, there exist transitive geodesics, form a class which is restricted in a topological as well as in a dynamical sense. For example, there is no known analytic two-dimensional manifold which is homeomorphic to a sphere or to a torus and has on it a transitive geodesic. The interrelation of the transitivity problem with the difficult problem of stability has been pointed out by Birkhoff [3].

By considering Fuchsian groups of the second kind we define manifolds M_{II}^n. As in the case of constant curvature, these are all open and the geodesic flow on any such manifold is not regionally transitive. Let a geodesic on M_{II}^n be *unstable* if, given any finite (compact) region on M_{II}^n, the points of the geodesic which lie in this region lie on a finite segment of the geodesic. Then the following theorem can be stated (cf. Hedlund [6]):

THEOREM 4.6. *Almost all the geodesics on any M_{II}^n are unstable.*

The method of proof is again based on horocycles.

Manifolds M_{II}^n have properties similar to the surfaces of negative curvature which Hadamard constructed (Hadamard [1]). A preliminary survey indicates that the methods used here can be applied to these Hadamard surfaces and that the perfect sets of geodesics constructed by Hadamard form sets of measure zero.

5. **Three-dimensional manifolds of constant negative curvature.** By assigning the metric

$$ds^2 = \frac{4(dx^2 + dy^2 + dz^2)}{c(1 - x^2 - y^2 - z^2)^2}$$

to the interior of the unit sphere $x^2+y^2+z^2=1$, and by identifying points which are congruent under the transformations of a properly discontinuous group which leaves this metric invariant, it is possible to define three-dimensional manifolds of constant negative curvature (cf. Tuller [1], where other references will be found). In contrast to the two-dimensional case, little seems to be known about the possible topological types of such manifolds. It is only recently that Löbell

[6] has constructed examples of this kind such that the manifolds are closed.

However, several of the methods which have been used in solving transitivity problems in connection with two-dimensional manifolds can be applied successfully to the case of three dimensions. Löbell [4] has derived theorems analogous to Theorems 3.1 and 3.2. Tuller [1] has shown that if almost all the geodesics are stable (where the definition of stable is analogous to that given in §4) almost all of them are transitive. Thus topometric transitivity holds on such manifolds which, in particular, include all those which are closed. Moreover, the work of Tuller indicates that under conditions similar to those in which the results hold in two dimensions, permanent regional and permanent topometric transitivity hold. E. Hopf (Zentralblatt für Mathematik, vol. 18, p. 273) states that his methods can be extended so as to prove metric transitivity in a large number of cases. The problem of mixture remains unsolved.

There appear to be no results concerning transitivity properties of the geodesics on three-dimensional manifolds which are not of constant curvature.

6. **Symbolic dynamics.** Symbolic methods have been used frequently in the derivation of transitivity properties of geodesics. These involve a characterization of the geodesic by an unending sequence of symbols called a symbolic trajectory. To a transitive geodesic corresponds a transitive symbolic trajectory; that is, one which contains every possible finite block, subject to certain rules of admissibility determined by the manifold under consideration. Conversely, to a transitive symbolic trajectory corresponds a transitive geodesic. With such a characterization available it is often a simple matter to construct a transitive symbolic trajectory, and thus prove the existence of a transitive geodesic.

In the case of the modular group (the interior of the unit circle being replaced by the upper half-plane with the Poincaré metric $(dx^2+dy^2)/y^2$) Artin [1] and Herglotz devised a symbolic characterization of the geodesics and with the aid of this characterization proved not only regional but topometric transitivity. Myrberg [1] independently derived similar results. Further analysis of this symbolic characterization enabled the author [2] to prove that metric transitivity holds in this case.

In the case of certain Fuchsian groups with symmetric fundamental region, Nielsen [2] employed symbolic methods to prove regional transitivity. The symbolic characterization of Nielsen (as given in [3]) was used by the author [1] in proving metric transitivity.

Koebe ([1], IV) has developed symbolisms in connection with general two-dimensional manifolds of constant negative curvature and used them to prove regional transitivity.

A symbolic analysis of the geodesics on certain surfaces of non-positive variable curvature and a proof of regional transitivity by means of this symbolism has been given by Birkhoff ([1], pp. 238–248). As shown by Birkhoff, much more than the existence of transitive geodesics can be inferred from the symbolic characterization. The symbolism enables one to dominate the problems concerning the qualitative behavior in the large of the geodesics.

The development of a symbolic theory apart from its dynamical significance has recently been begun by Morse and the author (cf. Morse [4]). This initial work includes an extensive analysis of transitive symbolic trajectories. The full scope of these symbolic methods in dynamics is yet to be determined.

Bibliography

E. Artin

1. *Ein mechanisches System mit quasiergodischen Bahnen*, Abhandlungen aus dem Mathematischen Seminar der Hamburgischen Universität, vol. 3 (1924), pp. 170–175.

G. D. Birkhoff

1. *Dynamical Systems*, American Mathematical Society Colloquium Publications, vol. 9, New York, 1927.
2. *Proof of a recurrence theorem for strongly transitive systems and proof of the ergodic theorem*, Proceedings of the National Academy of Sciences, vol. 17 (1931), pp. 650–660.
3. *Probability and physical systems*, this Bulletin, vol. 38 (1932), pp. 361–379.
4. (With P. A. Smith.) *Structure analysis of surface transformations*, Journal de Mathématiques Pures et Appliquées, (9), vol. 7 (1928), pp. 345–379.

L. R. Ford

1. *Automorphic Functions*, New York, 1929.

A. Grant

1. *Asymptotic transitivity on surfaces of variable negative curvature*, Bryn Mawr thesis, 1937

J. Hadamard

1. *Les surfaces à courbures opposées et leur lignes géodésiques*, Journal de Mathématiques Pures et Appliquées, (5), vol. 4 (1898), pp. 27–74.

G. A. Hedlund

1. *On the metrical transitivity of the geodesics on closed surfaces of constant negative curvature*, Annals of Mathematics, (2), vol. 35 (1934), pp. 787–808.
2. *A metrically transitive group defined by the modular group*, American Journal of Mathematics, vol. 57 (1935), pp. 668–678.
3. *Two-dimensional manifolds and transitivity*, Annals of Mathematics, (2), vol. 37 (1936), pp. 534–542.
4. *Fuchsian groups and transitive horocycles*, Duke Mathematical Journal, vol. 2 (1936), pp. 530–542.
5. See M. Morse [4].

6. *On the measure of geodesic types on surfaces of negative curvature*, Duke Mathematical Journal, vol. 5 (1939).
7. *Fuchsian groups and mixtures*, Annals of Mathematics, (2), vol. 40 (1939).

E. Hopf
1. *Fuchsian groups and ergodic theory*, Transactions of this Society, vol. 39 (1936), pp. 299–314.
2. *Ergodentheorie*, Ergebnisse der Mathematik und ihrer Grenzgebiete, vol. 5, 1937.

H. Hopf and W. Rinow
1. *Ueber den Begriff der vollständigen differentialgeometrischen Fläche*, Commentarii Mathematici Helvetici, vol. 3 (1931), pp. 209–225.

P. Koebe
1. *Riemannsche Mannigfaltigkeiten und nicht euklidische Raumformen*, Sitzungsberichte der Preussischen Akademie der Wissenschaften, I, 1927, pp. 164–196; II and III, 1928, pp. 345–442; IV, 1929, pp. 414–457; V and VI, 1930, pp. 304–364, 504–541; VII, 1931, pp. 506–534.

F. Löbell
1. *Die überall regulären unbegrenzten Flächen fester Krümmung*, Dissertation, Tübungen, 1927.
2. *Über die geodätischen Linien der Clifford-Kleinschen Flächen*, Mathematische Zeitschrift, vol. 30 (1929), pp. 572–607.
3. *Zur Frage der Struktur der geschlossenen geodätischen Linien in den offenen Clifford-Kleinschen Flächen mit positiver Charakteristik*, Journal für die reine und angewandte Mathematik, vol. 162 (1930), pp. 125–131.
4. *Einige Eigenschaften der Geraden in gewissen Clifford-Kleinschen Räumen*, Sitzungsberichte der Preussischen Akademie der Wissenschaften, 1930, pp. 556–569.
5. *Beispiele geschlossener dreidimensionaler Clifford-Kleinscher Räume negativer Krümmung*, Berichte über die Verhandlungen der sächsischen Akademie der Wissenschaften zu Leipzig, vol. 83 (1931), pp. 167–174.

Marston Morse
1. *Recurrent geodesics on a surface of negative curvature*, Transactions of this Society, vol. 22 (1921), pp. 84–100.
2. *A fundamental class of geodesics on any closed surface of genus greater than one*, Transactions of this Society, vol. 26 (1924), pp. 25–61.
3. *Instability and transitivity*, Journal de Mathématiques Pures et Appliquées, (9), vol. 14 (1935), pp. 49–71.
4. (With G. A. Hedlund) *Symbolic dynamics*, American Journal of Mathematics, vol. 60 (1938), pp. 815–866.

P. J. Myrberg
1. *Einige Anwendungen der Kettenbrüche in der Theorie der binären quadratischen Formen und der elliptischen Modulfunktionen*, Annales Academiae Scientiarum Fennicae, (A), vol. 23 (1925).
2. *Ein Satz über die fuchsschen Gruppen und seine Anwendung in der Funktionentheorie*, Annales Academiae Scientiarum Fennicae, (A), vol. 32 (1929).
3. *Ein Approximationssatz für die fuchsschen Gruppen*, Acta Mathematica, vol. 57 (1931), pp. 389–409.

J. von Neumann
1. *Einige Sätze über messbare Abbildungen*, Annals of Mathematics, (2), vol. 33 (1932), pp. 574–586.

J. Nielsen
1. *Om geodaetiske Linier i lukkede mangfoldigheder med konstant negativ Krumning*, Matematisk Tidsskrift, B, 1925, pp. 37–44.
2. *Zur Topologie der geschlossenen Flächen*, Mathematikerkongresses i Kopenhavn (1925), pp. 263–276.
3. *Untersuchungen zur Topologie der geschlossenen zweiseitigen Flächen*, Acta Mathematica, vol. 50 (1927), pp. 189–358.

H. Poincaré
1. *Sur les lignes géodésiques des surfaces convexes*, Transactions of this Society, vol. 6 (1905), pp. 237–274.

W. Rinow
See H. Hopf and W. Rinow.

W. Seidel
1. *On a metric property of Fuchsian groups*, Proceedings of the National Academy of Sciences, vol. 21 (1935), pp. 475–478.

P. A. Smith
See G. D. Birkhoff [4].

A. Tuller
1. *The measure of transitive geodesics on certain three-dimensional manifolds*, Duke Mathematical Journal, vol. 4 (1938), pp. 78–94.

The Institute for Advanced Study

REPRESENTATION OF NUMBERS IN TERNARY QUADRATIC FORMS

E. ROSENTHALL

We employ integral quaternions $t=t_0+t_1i_1+t_2i_2+t_3i_3$, where the coordinates t_i range over rational integers, while the i_1, i_2, i_3, satisfy the multiplication table

$$i_1^2 = i_2^2 = -2, \quad i_3^2 = -3, \quad i_2i_3 = 2i_1 - i_2 = (\overbrace{i_3i_2}),$$

$$i_3i_1 = -i_1 + 2i_2 = (\overbrace{i_1i_3}), \qquad i_1i_2 = -1 + i_3 = (\overbrace{i_2i_1}),$$

and $\bar{t}=t_0-t_1i_1-t_2i_2-t_3i_3$ is the conjugate to t. The norm $N(t)$ of t is $t\bar{t}=\bar{t}t=t_0^2+2t_1^2+2t_2^2+2t_1t_2+3t_3^2$. The norm of a product of two quaternions equals the product of their norms, and $\overline{vt}=\bar{t}\bar{v}$ for any two quaternions. The associative law $rs\cdot t=r\cdot st$ holds.

The quaternary quadratic $Q=t_0^2+2t_1^2+2t_2^2+2t_1t_2+3t_3^2$ has determinant 9, the g.c.d. of the literal coefficients of the adjoint to Q is 3, and the second concomitant of Q represents no residues 1 modulo 3, and as there is only one form of determinant 9 with these properties in Charve's table* of reduced quaternary quadratic forms, Q belongs to a genus of one class. Since Q represents 1 for two values of $t_0, \cdots, t_3$, we have,† a proper quaternion being defined as one having coprime coordinates, the following theorem:

THEOREM 1. *A proper quaternion $v=v_0+v_1i_1+v_2i_2+v_3i_3$ whose norm is divisible by a positive integer m prime to 6 has exactly two right-divisors (left-divisors) t and $-t$ of norm m.*

Every proper pure quaternion $s=s_1i_1+s_2i_2+s_3i_3$ of norm km^2 is of form $\bar{l}at$ where $N(a)=k$ and $N(t)=m$. For, $s=vt$ where $N(t)=m$ by Theorem 1; $\bar{s}=-s=\bar{l}\bar{v}$, and $\bar{l}$ is a left-divisor of s. Hence, since $N(v)=km$, $\bar{l}$ is a left-divisor of the proper quaternion v, $v=\bar{l}a$. Hence $s=\bar{l}at$, $N(a)=k$. Clearly a is pure since $\bar{l}\bar{a}t=-\bar{l}at$, $\bar{a}=-a$.

THEOREM 2. *Consider the equation $24n+1=x_1^2+2x_2^2+2x_3^2-2x_2x_3$. If $24n+1=m^2$, $(m>0)$, then all proper solutions are of type A if $m\equiv 1 \pmod 4$ but of type B if $m\equiv 3 \pmod 4$, where*

$$A\colon\ x_1\equiv\pm 1 \pmod{12}, \qquad B\colon\ x_1\equiv\pm 5 \pmod{12}.$$

---

* L. Charve, Comptes Rendus de l'Académie des Sciences, vol. 96 (1883), p. 773.

† G. Pall, *On the factorization of generalized quaternions*, submitted to the Duke Mathematical Journal.

If $24n+1$ is not a square, there are equally many solutions of each type A and B.

A proof for the case in which $24n+1=m^2$ follows. Consider

$$h = x_1^2 + 2x_2^2 + 2x_3^2 - 2x_2x_3. \tag{1}$$

Then

$$3h = 3x_1^2 + 2(-x_2 - x_3)^2 + 2(-x_2 - x_3)(2x_2 - x_3) + 2(2x_2 - x_3)^2.$$

Put $h=m^2$, and let $x=i_1(-x_2-x_3)+i_2(2x_2-x_3)+i_3x_1$ represent a solution (x_1, x_2, x_3) of (1); then all proper pure quaternions x are given by $x=\bar{t}at$, $N(t)=m$, $N(a)=3$, and from the latter condition we must have $a=\pm i_3$. Expanding $\bar{t}at$ gives $x_1=3t_3^2-2t_1^2-2t_2^2-2t_1t_2+t_0^2$ where only $a=i_3$ is considered since $a=-i_3$ merely changes the sign of x_1 which leaves A and B unaltered.

Thus $x_1\equiv m \pmod 4$. Since $(m, 3)=1$, $x_1\equiv 1 \pmod 3$; hence when $m\equiv 1 \pmod 4$, $x_1\equiv \pm 1 \pmod{12}$, a solution of type A, but if $m\equiv 3 \pmod 4$, then $x_1\equiv \pm 5 \pmod{12}$, a solution of type B.

The case for $h=24n+1$ not a square will now be considered. Let x be a representative solution of (1) under this condition. We can choose an odd prime p such that simultaneously

$$(-3h \mid p) = 1, \qquad p \equiv 11 \pmod{12}. \tag{2}$$

By the first equation in (2) we can choose x_0 so that $3x_0^2+h\equiv 0 \pmod p$. Then by Theorem 1, $3x_0+x$ has exactly two right-divisors $\pm t$ of norm p, say

$$3x_0 + x = ut, \qquad N(t) = p.$$

Then

$$txi\bar{t} = py, \qquad \text{where} \qquad y = tu - 3x_0, \tag{3}$$

and y represents another solution of (1). If t is replaced by $-t$, y is unchanged.

We shall prove that x and y are in opposite classes A and B in view of the second equation of (2), and as multiplication by p does not alter A or B, it will suffice to show that x and $txi\bar{t}$ are solutions of opposite types.

Setting $txi\bar{t}=i_1(-y_2-y_3)+i_2(2y_2-y_3)+i_3y_1$ and expanding gives

$$y_1 = x_1(t_0^2 + 3t_3^2 - 2t_2^2 - 2t_1^2 - 2t_1t_2) + x_2(6t_2t_3 + 4t_1t_0 + 2t_2t_0) + x_3(-6t_2t_3 - 6t_3t_1 + 2t_2t_0 - 2t_1t_0).$$

In (1), x_2 and x_3 are always even, and thus

$$y_1 \equiv 3x_1 \pmod 4, \qquad y_1 \equiv x_1 \pmod 3$$

as $N(\bar{t}) \equiv 11 \pmod{12}$. Hence y represents a solution of type opposite to x.

We can now establish the (1, 1) correspondence. We employ the preceding process for a fixed p, with x_0 for a solution of type A, but $-x_0$ for a solution of type B. Hence if our process carries x, a solution of type A, into y, then y is carried into x. For from (3)

$$3(-x_0) + y = (-\bar{u})\bar{t}, \qquad N(\bar{t}) = p.$$

Then, $\bar{t}yt = px$, $x = ut - 3x_0$. Further, two distinct solutions of one type cannot correspond to the same solution of the other type.

Application of other types of quaternions furnishes arithmetical proofs of the following additional results:

For representation of $24n+1$ in $x_1^2 + 3x_2^2 + 3x_3^2$ the A and B relations are

$$A: \begin{cases} x_1 \equiv \pm 1 \pmod{12}, & x_2 \text{ and } x_3 \equiv 0 \pmod 4, \\ x_1 \equiv \pm 5 \pmod{12}, & x_2 \text{ and } x_3 \equiv 2 \pmod 4, \end{cases}$$

$$B: \begin{cases} x_1 \equiv \pm 1 \pmod{12}, & x_2 \text{ and } x_3 \equiv 2 \pmod 4, \\ x_1 \equiv \pm 5 \pmod{12}, & x_2 \text{ and } x_3 \equiv 0 \pmod 4. \end{cases}$$

For $2(24n+1) = 3x_1^2 + x_2^2 + x_3^2$

A: (0; 1, 1), (0; 7, 7), (0; 11, 5), (4; 5, 5), (4; 11, 11), (4; 7, 1),
B: (4; 1, 1), (4; 7, 7), (4; 11, 5), (0; 5, 5), (0; 11, 11), (0; 7, 1),

where each triplet $(x_1; x_2, x_3)$ lists the least absolute residues $x_1 \pmod 8$, $x_2 \pmod{24}$, $x_3 \pmod{24}$ in a definite order.

For either form if $24n+1 = m^2$, $(m > 0)$, all solutions are of type A if $m \equiv 1 \pmod 6$, but of type B if $m \equiv 5 \pmod 6$. But there are equally many solutions of each type if $24n+1$ is not a square.

These results were proved in the writer's thesis at McGill University, 1938.

McGill University

ON APPROXIMATELY CONTINUOUS FUNCTIONS

ISAIAH MAXIMOFF

In a very interesting paper *Sur l'équation fonctionnelle* $g(x)=f\phi(x)$, S. Braun* established a series of theorems on the functional equation

$$g(x) = f[\phi(x)],$$

where $g(x)$ and $f(y)$ are given functions, and $\phi(x)$ is a function sought for. In this note, we consider the case for which $\phi(x)$ is an approximately continuous function.† A function $f(x)$ is said to be approximately continuous at x_0 if the density at x_0 of the set $E[f(x_0), \epsilon]$ of all points x such that $|f(x)-f(x_0)| < \epsilon$ is equal to 1, no matter what ϵ is.

Let $f(x)$ be a finite function of class 1 in $[0, 1]=[0\leqq x\leqq 1]$, and let $\{y_n\}$ be the sequence of all rational numbers y_n such that there are two points x_n' and x_n'' belonging to $[0, 1]$ and satisfying the condition $f(x_n') < y_n < f(x_n'')$. Let E_{y_n} (E^{y_n}), $(n=1, 2, 3, \cdots)$, denote the set of all points x such that $f(x)<y_n$ $(f(x)>y_n)$. If z is an irrational number, let E_z (E^z) denote the sum of all the sets E_{y_n} (E^{y_n}) such that $y_n<z$ $(y_n>z)$. We now prove the following theorem:

THEOREM 1. *A necessary and sufficient condition that a finite function* $\phi(x)$ *be approximately continuous in* $[0, 1]$ *is that there exist a system of perfect sets*

$$(\mathfrak{P}): \quad \mathfrak{P}_{y_r}^{n}, \quad \mathfrak{P}_{n}^{y_r}, \qquad r = 1, 2, 3, \cdots, n;\ n = 1, 2, 3, \cdots,$$

such that

(i) $E_{y_r}=\lim_{n=\infty}\mathfrak{P}_{y_r}^{n}$, $E^{y_r}=\lim_{n=\infty}\mathfrak{P}_{n}^{y_r}$, $\mathfrak{P}_{y_r}^{n}\subset\mathfrak{P}_{y_r}^{n+1}\subset E_{y_r}$, $\mathfrak{P}_{n}^{y_r}\subset\mathfrak{P}_{n+1}^{y_r}\subset E^{y_r}$;

(ii) *if* $y_r<y_s$ *and* M *is the greater of the integers* r, s, *every point of the set* $\mathfrak{P}_{y_r}^{n}$ $(\mathfrak{P}_{n}^{y_r})$ *is a density point of the set* $\mathfrak{P}_{y_s}^{n}$ $(\mathfrak{P}_{n}^{y_s})$ *for all* $n\geqq M$.

A point x of a set E will be called a density point in E if

$$\lim_{h=0}\left[\frac{1}{2h}\,\text{meas}\,[(x-h, x+h)E]\right]=1.$$

PROOF. Let x_0 be an arbitrary point in $[0, 1]$, and let $f(x_0)=y_0$.

* Fundamenta Mathematicae, vol. 28 (1937), pp. 294–302.

† A. Denjoy, *Sur les fonctions dérivées sommables*, Bulletin de la Société Mathématique de France, vol. 43 (1915), pp. 161–247, especially p. 165.

We suppose that y_0 is not the least upper bound (greatest lower bound) of f in $[0, 1]$. Then, if η is an arbitrary, positive, sufficiently small number, we have integers k_1, k, s, s_1 such that $y_0-\eta<y_{k_1}<y_k<y_0<y_s<y_{s_1}<y_0+\eta$.

The point x_0 belongs to the set $E_{y_s}\cdot E_{y_{s_1}}$ as well as to the set $E^{y_k}\cdot E^{y_{k_1}}$. Consequently, x_0 belongs to the set $\mathfrak{P}^n_{y_s}\cdot\mathfrak{P}^n_{y_{s_1}}$ as well as to the set $\mathfrak{P}^{y_k}_n\cdot\mathfrak{P}^{y_{k_1}}_n$ for sufficiently large n. Thence x_0 is a density point in $\mathfrak{P}^n_{y_{s_1}}$ as well as in $\mathfrak{P}^{y_{k_1}}_n$; therefore x_0 is a density point in $\mathfrak{P}^n_{y_{s_1}}\cdot\mathfrak{P}^{y_{k_1}}_n$. But if $E[f(x_0), \eta]$ is the set of all points x satisfying the condition $|f(x)-f(x_0)|\leqq\eta$, we have

$$\mathfrak{P}^n_{y_{s_1}}\cdot\mathfrak{P}^{y_{k_1}}_n \subset E_{y_{s_1}}\cdot E^{y_{k_1}} \subset E[f(x_0), \eta];$$

so that x_0 is a density point in $E[f(x_0), \eta]$.

We now suppose that y_0 is the least upper bound (greatest lower bound) of f in $[0, 1]$. Then, if η is an arbitrary, positive, sufficiently small number, we have integers k_1, k (s, s_1) such that

$$y_0 - \eta < y_{k_1} < y_k < y_0 \quad (y_0 < y_s < y_{s_1} < y_0 + \eta).$$

The point x_0 belongs to the set

$$E^{y_k}\cdot E^{y_{k_1}} \quad (E_{y_s}\cdot E_{y_{s_1}}).$$

Consequently, x_0 belongs to the set

$$\mathfrak{P}^{y_k}_n\cdot\mathfrak{P}^{y_{k_1}}_n \quad (\mathfrak{P}^n_{y_s}\cdot\mathfrak{P}^n_{y_{s_1}})$$

for sufficiently large n. Thence x_0 is a density point in $\mathfrak{P}^{y_{k_1}}_n$ $(\mathfrak{P}^n_{y_{s_1}})$, and x_0 is therefore a density point in

$$E^{y_{k_1}} \subset E[f(x_0), \eta] \quad (E_{y_{s_1}} \subset E[f(x_0), \eta]),$$

so that x_0 is a density point in $E[f(x_0), \eta]$, and the sufficiency of the condition is established. To prove that the condition is necessary, we shall assume that $f(x)$ is an approximately continuous function. In virtue of a Theorem of A. Denjoy,* in this case $f(x)$ is a function of class 1. It follows that each of the sets E_{y_r}, E^{y_r}, $(r=1, 2, 3, \cdots)$, is the sum of an enumerable infinity of perfect sets and of an enumerable set N of points $x_1, x_2, x_3, \cdots$ of $[0, 1]$. Let E denote an arbitrary one of the sets E_{y_n}, E^{y_n}, $(n=1, 2, 3, \cdots)$. Since $f(x)$ is approximately continuous, we can find for every point x_n a perfect set $\mathfrak{P}(x_n)$ such that

* Loc. cit., p. 181.

(i) x_n is a density point of $\mathfrak{P}(x_n)$;
(ii) $f(x)$ is continuous at the point x_0 over $\mathfrak{P}(x_n)$ relative to $\mathfrak{P}(x_n)$;
(iii) $\mathfrak{P}(x_n)$ is contained in E.

In this case $\mathfrak{P}(x_n)$ will be called the perfect and dense *road* at the point x_n. It is clear that E is the sum of perfect sets and of sets $\mathfrak{P}(x_n)$, $(n=1, 2, 3, 4, \cdots)$. Since $\mathfrak{P}(x_n)$ is perfect, E is the enumerable sum of perfect sets, and we may also write

$$E = \lim_{n=\infty} P_n,$$

where P_n is a perfect set such that $P_n \subset E$.

We have also

$$E_{y_r} = \lim_{n=\infty} P^n_{y_r}, \quad E_{y_s} = \lim_{n=\infty} P^n_{y_s}, \quad E^{y_r} = \lim_{n=\infty} P^{y_r}_n, \quad E^{y_s} = \lim_{n=\infty} P^{y_s}_n,$$

where $P^n_{y_r}$, $P^n_{y_s}$, $P^{y_r}_n$, $P^{y_s}_n$, are perfect sets such that

$$P^n_{y_r} \subset E_{y_r}, \quad P^n_{y_s} \subset E_{y_s}, \quad P^{y_r}_n \subset E^{y_r}, \quad P^{y_s}_n \subset E^{y_s}.$$

Hence on putting $\mathfrak{P}^n_{y_s} = P^n_{y_s} + P^n_{y_r}$, $\mathfrak{P}^{y_r}_n = P^{y_r}_n + P^{y_s}_n$, we have

$$E_{y_s} = \lim_{n=\infty} \mathfrak{P}^n_{y_s}, \qquad E^{y_r} = \lim_{n=\infty} \mathfrak{P}^{y_r}_n.$$

It is easily seen that

$$\mathfrak{P}^n_{y_r} \subset \mathfrak{P}^n_{y_s}, \qquad \mathfrak{P}^{y_s}_n \subset \mathfrak{P}^{y_r}_n.$$

Let (a, b) be any contiguous interval of the set $\mathfrak{P}^n_{y_r}$ $(\mathfrak{P}^{y_s}_n)$. We shall construct a perfect set Q_{ab} (R_{ab}) such that

(i) $Q_{ab} \subset E_{y_r}$, $R_{ab} \subset E^{y_s}$;
(ii) every point of the set $\mathfrak{P}^n_{y_r}$ $(\mathfrak{P}^{y_s}_n)$ is the density point of the set

$$\Omega^n_{y_r} = \mathfrak{P}^n_{y_r} + \sum_{ab} Q_{ab} \left(\Omega^{y_s}_n = \mathfrak{P}^{y_s}_n + \sum_{ab} R_{ab} \right).^*$$

Now we can adjoin the set $\Omega^n_{y_r}$ $(\Omega^{y_r}_n)$ to the set $\mathfrak{P}^n_{y_s}$ $(\mathfrak{P}^{y_r}_n)$. Thus, without loss of generality, we may assume that every point of the set $\mathfrak{P}^n_{y_r}$ $(\mathfrak{P}^{y_r}_n)$ is a density point of the set $\mathfrak{P}^n_{y_s}$ $(\mathfrak{P}^{y_r}_n)$ for all $n \geqq M$.

We now turn to the functional equation

$$g(x) = f[\phi(x)].$$

Consider the sequence of all rational numbers (y): $y_1, y_2, y_3, \cdots$.

* See V. Bogomoloff, *Sur une classe des fonctions asymptotiquement continues*, Recueil Mathématique, Moscow, vol. 32 (1924), pp. 152–171.

Let $\mathfrak{E}_{y_r}$ be the set of all points (x, y) such that

$$g(x) = f(y), \qquad y < y_r, 0 \leqq x \leqq 1,$$

and let $\mathfrak{E}^{y_r}$ be the set of all points (x, y) such that

$$g(x) = f(y), \qquad y > y_r, 0 \leqq x \leqq 1.$$

Denote by $\mathfrak{P}\mathfrak{E}$ the orthogonal projection of a set $\mathfrak{E}$ on the x axis.

Suppose now that M is the greater of the integers r, s, and assume $y_r < y_s$.

THEOREM 2. *A necessary and sufficient condition that there exist a finite function $\phi(x)$ of class 1 having the property of Darboux and satisfying the equation $g(x) = f[\phi(x)]$ is that there exist a sequence of perfect sets*

$$(\mathfrak{P}): \quad \mathfrak{P}^n_{y_r}, \quad \mathfrak{P}^{y_r}_n, \qquad r = 1, 2, 3, \cdots, n; n = 1, 2, 3, \cdots,$$

such that

(i) *the sets satisfy*

$$\lim_{n=\infty} \mathfrak{P}^n_{y_r} \subset \mathfrak{P}\mathfrak{E}_{y_r}, \quad \lim_{n=\infty} \mathfrak{P}^{y_r}_n \subset \mathfrak{P}\mathfrak{E}^{y_r}, \quad \mathfrak{P}^n_{y_r} \subset \mathfrak{P}^{n+1}_{y_r} \subset \mathfrak{P}\mathfrak{E}_{y_r}, \quad \mathfrak{P}^{y_r}_n \subset \mathfrak{P}^{y_r}_{n+1} \subset \mathfrak{P}\mathfrak{E}^{y_r};$$

(ii) *every point of the set $\mathfrak{P}^n_{y_r}$ ($\mathfrak{P}^{y_r}_n$) is a density point of the set $\mathfrak{P}^n_{y_s}$ ($\mathfrak{P}^{y_r}_n$) for all $n \geqq M$;*

(iii) $\lim_{n=\infty} \sum_{r=1}^{r=n} \mathfrak{P}^n_{y_r} = \lim_{n=\infty} \sum_{r=1}^{r=n} \mathfrak{P}^{y_r}_n = [0 \leqq x \leqq 1]$.

PROOF. To prove that the condition is necessary, we shall assume that there exists a finite function $\phi(x)$ of class 1 having the property of Darboux and satisfying the equation $g(x) = f[\phi(x)]$. Let $\overline{\mathfrak{E}}_{y_r}$ ($\overline{\mathfrak{E}}^{y_r}$) be the set of all points $[x, \phi(x)]$, where x belongs to the set E_{y_r} (E^{y_r}). It is clear that

$$\mathfrak{P}\overline{\mathfrak{E}}_{y_r} \subset \mathfrak{P}\mathfrak{E}_{y_r}, \qquad \mathfrak{P}\overline{\mathfrak{E}}^{y_r} \subset \mathfrak{P}\mathfrak{E}^{y_r}.$$

Since $\mathfrak{P}\overline{\mathfrak{E}}_{y_r} = E_{y_r}$, $\mathfrak{P}\overline{\mathfrak{E}}^{y_r} = E^{y_r}$, we conclude at once that $E_{y_r} \subset \mathfrak{P}\mathfrak{E}_{y_r}$, $E^{y_r} \subset \mathfrak{P}\mathfrak{E}^{y_r}$. The sum of all sets E_{y_r} (E^{y_r}) is equal to the segment $[0 \leqq x \leqq 1]$, hence

$$\text{(A)} \qquad \sum_{r=1}^{r=\infty} \mathfrak{P}\mathfrak{E}_{y_r} = \sum_{r=1}^{r=\infty} \mathfrak{P}\mathfrak{E}^{y_r} = [0 \leqq x \leqq 1].$$

In virtue of Theorem 1 for the function $\phi(x)$, there exists a sequence of perfect sets

$$(\mathfrak{P}): \quad \mathfrak{P}^n_{y_r}, \quad \mathfrak{P}^{y_r}_n$$

satisfying the conditions of Theorem 1. It is readily seen that

$$\text{(B)} \qquad \mathfrak{P}^{n}_{y_r} \subset \mathfrak{P}\mathfrak{E}_{y_r}, \qquad \mathfrak{P}^{y_r}_{n} \subset \mathfrak{P}\mathfrak{E}^{y_r}.$$

We have just seen that the condition is necessary; let us next show that it is sufficient. For this purpose, we shall assume that there exists a system of perfect sets

$$(\mathfrak{P}): \qquad \mathfrak{P}^{n}_{y_r}, \quad \mathfrak{P}^{y_r}_{n}, \qquad r = 1, 2, 3, \cdots, n; \; n = 1, 2, 3, \cdots,$$

satisfying the following conditions:

(i) $\mathfrak{P}^{n}_{y_r} \subset \mathfrak{P}^{n+1}_{y_r} \subset \mathfrak{P}\mathfrak{E}_{y_r}$, $\mathfrak{P}^{y_r}_{n} \subset \mathfrak{P}^{y_r}_{n+1} \subset \mathfrak{P}\mathfrak{E}^{y_r}$.

(ii) Every point of the set $\mathfrak{P}^{n}_{y_r}$ ($\mathfrak{P}^{y_r}_{n}$) is a density point of the set $\mathfrak{P}^{n}_{y_r}$ ($\mathfrak{P}^{y_r}_{n}$) for all $n \geqq M$.

(iii) $\lim_{n=\infty} \sum_{r=1}^{r=n} \mathfrak{P}^{n}_{y_r} = \lim_{n=\infty} \sum_{r=1}^{r=n} \mathfrak{P}^{y_r}_{n} = [0 \leqq x \leqq 1]$.

It follows that there exists a finite function $\phi(x)$ of class 1 having the property of Darboux and satisfying the condition $g(x) = f[\phi(x)]$. In order to find the function $\phi(x)$, we shall proceed as follows. We first effect the uniformization* of the set $\mathfrak{E}_{y_1}$ relative to the y axis over the set $\mathfrak{P}^{1}_{y_1}$; thus the value of the function $\phi(x)$ is determined at each point x of $\mathfrak{P}^{1}_{y_1}$. We next effect the uniformization of the set $\mathfrak{E}^{y_1}$ over the set

$$R_1 = \mathfrak{P}^{y_1}_{1} - \mathfrak{P}^{y_1}_{1} \cdot \mathfrak{P}^{1}_{y_1}.$$

We set $U_1 = \mathfrak{P}^{y_1}_{1} + \mathfrak{P}^{1}_{y_1}$. We then effect the uniformization of the set $\mathfrak{E}_{y_1}$ over the set

$$R_2 = \mathfrak{P}^{2}_{y_1} - U_1 \cdot \mathfrak{P}^{2}_{y_1}.$$

Set $U_2 = U_1 + \mathfrak{P}^{2}_{y_1}$. We now effect the uniformization of the set $\mathfrak{E}_{y_2}$ over the set

$$R_3 = \mathfrak{P}^{2}_{y_2} - U_2 \cdot \mathfrak{P}^{2}_{y_2}$$

and set $U_3 = U_2 + \mathfrak{P}^{2}_{y_2}$. We next effect the uniformization of the set $\mathfrak{E}_{y_1}$ over the set

$$R_4 = \mathfrak{P}^{3}_{y_1} - U_3 \cdot \mathfrak{P}^{3}_{y_1},$$

and so on.

We carry out this process until the function $\phi(x)$ is completely determined.

TCHEBOKSARY, U.S.S.R.

* See N. Lusin, *Sur le problème de M. Jacques Hadamard d'uniformisation des ensembles*, Comptes Rendus de l'Académie, Paris, vol. 189 (1930), p. 349.

ON FERMAT'S SIMPLE THEOREM

JACK CHERNICK

1. **Introduction.** Fermat's simple theorem may be stated as follows: *If a is any integer prime to m, and if m is prime, then*

$$a^{m-1} \equiv 1 \pmod{m}. \tag{1}$$

The question naturally arises, "Do there exist composite integers for which the same congruence holds?" For particular values of a the existence of such numbers has long been established.* In 1910, R. D. Carmichael† treated the congruence (1) in the stricter sense indicated. He established several criteria which may be condensed into the following theorem:

THEOREM 1. *Fermat's theorem holds for composite integers if and only if m may be expressed as a product of distinct odd primes $p_1, \cdots, p_n$, $(n>2)$, and $m-1\equiv 0 \pmod{p_i-1}$ where i ranges from 1 to n.*

Carmichael listed several such m with $n=3$ and one with $n=4$. Many others have since been found by P. Poulet.‡ It is our purpose to continue the study of these numbers in the present paper.

Fermat's theorem is sometimes stated thus: *If m is any prime and a any integer, then*

$$a^m \equiv a \pmod{m}. \tag{2}$$

The congruences (1) and (2) are likewise equivalent if m is composite, as is easily shown by the use of Theorem 1.

Despite the apparent promise of Fermat's theorem of yielding a complete and practical test for primes, no modification of it has as yet achieved this goal. However, the recent work of D. H. Lehmer,§ based upon a list of solutions of (2) for $a=2$, now provides such a test for integers in the range 10^7 to 10^8.

2. **Proof of Theorem** 1. We present a short, independent proof of Theorem 1. Let m be a composite number for which (1) holds. First, suppose $m=2^v$, $(v>1)$. But $a^{2^{v-1}}\equiv 1 \pmod{2^v}$ will not hold for

* Dickson, *History of the Theory of Numbers*, vol. 1, pp. 92–95.

† This Bulletin, vol. 16 (1910), pp. 232–238; also American Mathematical Monthly, vol. 19 (1912), pp. 22–27.

‡ D. H. Lehmer informs us that all m's under $5\cdot 10^7$ and all, with $n=3$, under 10^8 have been tabulated by Poulet.

§ American Mathematical Monthly, vol. 43 (1936), pp. 347–354.

$a=3$ and $v\geqq 2$. Therefore m contains at least one odd prime factor.

Next, let $m=rp^v$ where v is the highest power of any odd prime contained in m. Let ω be a primitive root of p^v. Since ω is prime to p, the arithmetical progression ω, $\omega+p^v$, $\omega+2p^v$, $\cdots$ includes an infinitude of primes. Select s sufficiently large so that $x=\omega+sp^v$ is a prime greater than m. Then x is prime to m, and by (1),

$$x^{m-1}\equiv\omega^{m-1}\equiv 1 \pmod{p^v}.$$

Since ω is a primitive root of p^v, $m-1\equiv 0 \pmod{p^v-p^{v-1}}$. But $m-1=rp^v-1$ is prime to p. Hence $v=1$ only and

$$m-1\equiv 0 \pmod{p-1}. \tag{3}$$

Also $p-1$ is even. Hence m is odd.

It remains to show that $n>2$. Else write $m=p_1p_2$, $(p_1>p_2)$. Then by (3),

$$p_1p_2-1\equiv p_2-1\equiv 0 \pmod{p_1-1};$$

or $p_2\geqq p_1$, a contradiction.

This completes the proof that the conditions given in Theorem 1 are necessary. Conversely, when m satisfies the stated conditions, the congruence (1) obviously follows.

We shall find it convenient henceforth to denote by F_n any composite integer of n prime factors for which Fermat's theorem holds.

3. **Properties of F_3.** A. *Theorem* 2. We shall prove the following theorem:

THEOREM 2. *Every F_3 is of the form $(2r_1h+1)(2r_2h+1)(2r_3h+1)$ where the r's are relatively prime in pairs.*

Let $F_3=p_1p_2p_3$. Set $p_i=r_ik+1$, where k is the g.c.f. of p_i-1, i running from 1 to 3. Then by Theorem 1, we have the congruential conditions

$$(r_1k+1)(r_2k+1)(r_3k+1)\equiv 1 \pmod{kr_i};$$

or by simplifying,

$$k(r_1r_2+r_1r_3+r_2r_3)+r_1+r_2+r_3\equiv 0 \pmod{r_i}. \tag{4}$$

The r's are relatively prime in pairs; for by (4), if any two have a common factor, so does the third, contrary to hypothesis. Since k must be even, we obtain Theorem 2.

Now (4) is replaceable by the single condition

$$k(r_1r_2+r_1r_3+r_2r_3)+r_1+r_2+r_3\equiv 0 \pmod{r_1r_2r_3}. \tag{5}$$

The latter congruence is linear in k. Moreover, the coefficient of k is prime to the modulus. Hence its general solution is given by

$$k = Mr_1r_2r_3 - (r_1 + r_2 + r_3)(r_1r_2 + r_1r_3 + r_2r_3)^a,$$

where $a = \phi(r_1)\phi(r_2)\phi(r_3) - 1$.

B. *Universal forms.* For given r_i the solution of (5) affords a one-parameter expression for F_3. Thus $(r_1, r_2, r_3) = (1, 2, 3)$ gives $k = 6M$, whence

$$U_3 = (6M + 1)(12M + 1)(18M + 1)$$

yields an F_3 for every M for which the quantities in parentheses are prime. A few examples are

$7 \cdot 13 \cdot 19$,	$271 \cdot 541 \cdot 811$,	$337 \cdot 673 \cdot 1009$,
$37 \cdot 73 \cdot 109$,	$307 \cdot 613 \cdot 919$,	$601 \cdot 1201 \cdot 1801$,
$211 \cdot 421 \cdot 631$,	$331 \cdot 661 \cdot 991$,	$727 \cdot 1453 \cdot 2179$.

Similarly, $(r_1, r_2, r_3) = (1, 2, 5)$, $(1, 3, 8)$, and $(2, 3, 5)$ yield, respectively, the forms $U_3 = (10M+7)(20M+13)(50M+31)$, $(24M+13) \cdot (72M+37)(192M+97)$, and $(60M+41)(90M+61)(150M+101)$.

We shall call these forms universal. More precisely, the product U_n of n odd distinct linear factors a_iM+b_i, $(n \geqq 3)$, will be termed universal if it satisfies the set of congruences $U_n \equiv 1 \pmod{a_iM+b_i-1}$, where i ranges from 1 to n, for every integral value of M. The presence of these forms makes it easy to conjecture but no less difficult to prove the existence of an infinitude of F_n. The question whether such forms represent an infinitude of sets of primes has already been raised by L. E. Dickson.*

4. **Properties of F_n, $(n > 3)$.** When $n > 3$, similar results may be derived. If we write any F_n in the form $(r_1k+1) \cdots (r_nk+1)$, where k is the g.c.f. of $p_i - 1$, i ranging from 1 to n, it can be shown that the r's are relatively prime in sets of $n-1$. But for given r_i, the congruence of Theorem 1 is no longer readily nor necessarily solvable. Thus if $n = 4$, (4) is replaced by the quadratic congruences

$$k^2(\sum r_1r_2r_3) + k(\sum r_1r_2) + \sum r_1 \equiv 0 \pmod{r_i}, \tag{6}$$

i from 1 to 4. Let us limit the r's by $\prod_1^4 r_1 \leqq 100$. Most of the possible cases are then eliminated by the theory of quadratic residues. By solving (6) generally for the six cases that remain, there results from each one or more universal forms, as exhibited in the following table:

* Messenger of Mathematics, vol. 33 (1904), pp. 155–161.

r_i	U_4
1, 2, 3, 6	$(6M+1)(12M+1)(18M+1)(36M+1)$
1, 2, 3, 12	$(12M+7)(24M+13)(36M+19)(144M+73)$
1, 2, 4, 7	$(28M+15)(56M+29)(112M+57)(196M+99)$
1, 2, 4, 11	$(44M+15)(88M+29)(176M+57)(484M+155)$
1, 2, 4, 11	$(44M+43)(88M+85)(176M+169)(484M+463)$
1, 2, 5, 10	$(10M+7)(20M+13)(50M+31)(100M+61)$
1, 3, 4, 6	$(12M+11)(36M+31)(48M+41)(72M+61)$

When $n>4$, congruences of higher degree than (6) result. To surmount this difficulty, we have the following theorem:

THEOREM 3. *Let $p_1p_2\cdots p_n$ be an F_n. Define k_1 as the g.c.f. of p_i-1, $r_i=(p_i-1)/k_1$, and R as the l.c.m. of r_i, i ranging from 1 to n. Then $U_n=\prod_1^n(r_iRM+p_i)$ is an universal form, with the proviso that if the r's are all odd, M be replaced by $2M$.*

By Theorem 1, $k=k_1$ is a solution of the congruence

$$\left[\prod_1^n (r_ik+1)-1\right]/k \equiv 0 \pmod{R}. \tag{7}$$

Hence any $k\equiv k_1 \pmod R$ is a solution of (7). Let $k=MR+k_1$. On substituting this for k_1, we obtain the form of Theorem 3. By (7), this form satisfies the congruence required for universality. Any factor r_iRM+p_i of this form is odd since RM is even and p_i is odd. No two factors r_iRM+p_i, r_jRM+p_j are equal, for if $p_i>p_j$, by definition $r_i>r_j$. Hence the form is universal.

Theorem 3 enables us to derive universal forms from given F_n. A method of obtaining such F_n in certain cases from known F_{n-1} is now shown by the next theorem:

THEOREM 4. *Let $F_{n-1}=p_1p_2\cdots p_{n-1}$, q the l.c.m. of p_i-1, i from 1 to $n-1$, and $r=(F_{n-1}-1)/q$. If $p_n=q\pi+1$, where π is any divisor of r and p_n is a prime distinct from p_i, then $p_1p_2\cdots p_n$ is an F_n.*

By Theorem 1, it suffices that $F_n\equiv 1 \pmod{p_i-1}$, i ranging from 1 to n. Now $F_n\equiv F_{n-1}p_n\equiv p_n\equiv 1 \pmod q$. Hence it remains to show that $F_n\equiv 1 \pmod{p_n-1}$. But $F_n\equiv F_{n-1}p_n\equiv F_{n-1}\equiv 1 \pmod{p_n-1}$, whence Theorem 4 follows.

As an example, take $F_3=7\cdot 13\cdot 19$. Then $q=36$, $r=48$, $p_4=37$, 73, 109, 433, or 577. By repeated application of Theorem 4, we find the interesting series

$F_2 = 5 \cdot 17 \cdot 29,\ 7 \cdot 13 \cdot 31,$
$F_4 = 5 \cdot 17 \cdot 29 \cdot 113,\ 7 \cdot 13 \cdot 31 \cdot 61,$
$F_5 = 5 \cdot 17 \cdot 29 \cdot 113 \cdot 337,\ 7 \cdot 13 \cdot 31 \cdot 61 \cdot 181,$
$F_6 = 5 \cdot 17 \cdot 29 \cdot 113 \cdot 337 \cdot 673,\ 7 \cdot 13 \cdot 31 \cdot 61 \cdot 181 \cdot 541,$
$F_7 = 5 \cdot 17 \cdot 29 \cdot 113 \cdot 337 \cdot 673 \cdot 2689,\ 7 \cdot 13 \cdot 31 \cdot 61 \cdot 181 \cdot 541 \cdot 2161.$

The process may be continued to the limits of present-day factor tables. With the aid of Theorem 3, we can use these solutions to derive universal forms. Thus from the last of these F_n, we get

$$U_7 = (360M + 7)(720M + 13)(1800M + 31)(3600M + 61) \cdot (10800M + 181)(32400M + 541)(129600M + 2161).$$

5. **The existence of an U_n for any $n > 3$.** Theorem 4 is readily applied to U_n in place of F_n by merely omitting the condition that the p's be prime. For instance, consider $U_3 = (6M+1)(12M+1)(18M+1)$. Here $q = 36M$. Taking $\pi = 1$, we obtain

$$U_4 = (6M + 1)(12M + 1)(18M + 1)(36M + 1).$$

Similarly, we find

$$U_5 = (6M + 1)(12M + 1)(18M + 1)(36M + 1)(72M + 1),$$

provided $M \equiv 0 \pmod 2$; and

$$U_6 = (6M + 1)(12M + 1)(18M + 1)(36M + 1)(72M + 1)(144M + 1),$$

if $M \equiv 0 \pmod 4$.

This suggests the possibility of an unending series of such forms. Indeed, suppose $M = 2^{n-4}M_1$, and let

$$(8) \quad U_n = (6M+1)(12M+1)(18M+1)(2^2 \cdot 9M+1) \cdots (2^{n-2} \cdot 9M+1)$$

be an universal form. Then

$$U_{n+1} = (6M + 1)(12M + 1)(18M + 1)(2^2 \cdot 9M + 1) \cdots (2^{n-1} \cdot 9M + 1)$$

is universal if

$$(9) \qquad U_n \equiv 1 \pmod{2^{n-1} \cdot 9M},$$

since, by Theorem 4, $q = 2^{n-2} \cdot 9M$ and we may thus take $\pi = 2$. By Theorem 1, we already know that (9) holds for the modulus $2^{n-2} \cdot 9M$.

When U_n is expanded in terms of M, (9) becomes

$$(10) \quad 1 + M(6+12+18+36+ \cdots + 2^{n-2} \cdot 9) + KM^2 \equiv 1 \pmod{2^{n-1} \cdot 9M},$$

where K is a polynomial in M. The second term in the left-hand member sums to $2^{n-1} \cdot 9M$. Hence (10) reduces to

$$KM \equiv 0 \pmod{2^{n-1} \cdot 9}. \tag{11}$$

Conversely (11) implies (9). Since (9) holds for the modulus $2^{n-2} \cdot 9M$, it follows similarly that (11) holds for the modulus $2^{n-2} \cdot 9$ with $M = 2^{n-4}M_1$. Hence (11) will be true for the given modulus if $M = 2^{n-3}M_1$. This supplies a proof by induction that (8) is a universal form for every $n \geqq 4$.

If, in addition,* M is divisible by every prime p where $3 < p \leqq n$, we satisfy the necessary condition given by Dickson† for the form (8) to represent at least one set of n primes. The proof of the sufficiency of this condition still remains a challenge to the ingenuity of number theorists.

NEW YORK, N. Y.

RINGS AS GROUPS WITH OPERATORS

C. J. EVERETT, JR.

1. **Introduction.** A module M $(0, a, b, \cdots)$ is a commutative group, additively written. Every correspondence of M onto itself, or part of itself, such that $a \to a'$, $b \to b'$ implies $a+b \to a'+b'$ defines an *endomorphism* of M. An endomorphism may be regarded as an operator θ on M subject to the postulates (i) $\theta a = a'$ is uniquely defined as an element of M, (ii) $\theta(a+b) = \theta a + \theta b$, $(a, b \in M)$. In particular, there exist a null operator 0 $(0M = 0)$ and a unit operator ϵ $(\epsilon a = a, a \in M)$. Designate by Ω_M the set of all such operators, $0, \epsilon, \alpha, \beta, \cdots$. It is well known that if operations of $\oplus$ and $\odot$ be defined in Ω_M by $(\theta+\eta)a = \theta a + \eta a$ and $(\theta\eta)a = \theta(\eta a)$, $(a \in M)$, Ω_M forms a ring with unit element ϵ (*endomorphism ring* of M).‡ The equation $\theta = \eta$ means $\theta a = \eta a$ (all $a \in M$). A ring $R(M)$ is called a ring over M in case M is the additive group of $R(M)$. Correspondence of a set P onto a set Q (many-one) is written $P \sim Q$; if specifically one-one, $P \cong Q$. Corresponding operations in P, Q preserved under the map are indicated in parentheses; for example, $P \sim Q$ $(+)$. If a set T has the property that TP is defined in P, TQ in Q, and if, under a correspondence $P \sim Q$, $p \to q$ implies $tp \to tq$ $(t \in T, p \in P, q \in Q)$, we write $P \sim Q$ (T) (T-operator correspondence). If R is a ring, the two-sided ideal N of elements z of R such that $zr = 0$ (all $r \in R$), is called the left annulling ideal of R.

---

* For example, replace $6M$ in (8) by $2^w n!M$, $(w \geqq n-3)$.

† Loc. cit., p. 156.

‡ van der Waerden, *Moderne Algebra*, vol. 1, 2d edition, p. 146.

2. **Fundamental theorems.** We prove first the following theorem:

THEOREM 1. *If $R(M)$ is a ring over M, there exists in Ω_M a subring Γ such that*

$$R(M) \sim \Gamma \quad (\oplus, \odot; \Gamma),$$

*this correspondence being one-one if and only if $N=(0)$ for $R(M)$.**

For $R(M)$ consists of the elements of M on which a multiplication has been defined so that (i) $ab \in M$, (ii) $a(b+c)=ab+ac$, (iii) $(a+b)c=ac+bc$, (iv) $(ab)c=a(bc)$. By (i), every a of M defines a map of M into M which by (ii) is an endomorphism. Hence to every a of M corresponds an operator α of Ω_M. Let Γ be the set of all such α, whence $R(M)\sim\Gamma$, where $a\rightarrow\alpha$ is defined by $ag=\alpha g$ (all $g \in M$). We have that $a+b\rightarrow\alpha+\beta$, $ab\rightarrow\alpha\beta$ and $\gamma a\rightarrow\gamma\alpha$ from the following:

$$\begin{aligned} (a+b)h &= ah+bh = \alpha h+\beta h = (\alpha+\beta)h, \\ (ab)h &= a(bh) = a(\beta h) = \alpha(\beta h) = (\alpha\beta)h, \\ (\gamma a)h &= (ga)h = g(ah) = (\gamma\alpha)h, \qquad \text{all } h \in M. \end{aligned}$$

Since, under the correspondence, $N\rightarrow 0$, proof of the theorem is complete.

THEOREM 2. *If in Ω_M there exists a subring Γ such that $M\sim\Gamma\ (\oplus;\Gamma)$ then there exists a ring $R(M)$ over M such that*

$$R(M) \sim \Gamma \quad (\oplus, \odot; \Gamma).$$

We define $ab=\alpha b$. Then

$$a(b+c) = \alpha(b+c) = \alpha b + \alpha c = ab + ac, \tag{1}$$

$$(a+b)c = (\alpha+\beta)c = \alpha c + \beta c = ac + bc, \tag{2}$$

$$(ab)c = (\alpha b)c = (\alpha\beta)c = \alpha(\beta c) = \alpha(bc) = a(bc), \tag{3}$$

and M with this multiplication is a ring $R(M)$. Since $ab=\alpha b\rightarrow\alpha\beta$, the theorem follows.

COROLLARY. *If $M\sim\Gamma\ (\oplus)$, Γ a submodule of Ω_M, there exists a (non-associative) ring $R^*(M)$ over M, where ab is defined as αb, $(a\rightarrow\alpha)$.*

The relation between associativity of $R(M)$ and the Γ-operator character of the correspondence seems to indicate a point of departure for the study of rings with associativity not assumed.

* In case $N\neq(0)$, there exists a ring $R_1\supset R$ for which $N_1=(0)$; thus R is always isomorphic with a subring of the endomorphism ring of some module. See, for example, A. A. Albert, *Modern Higher Algebra*, University of Chicago Press, 1937, p. 22, Theorem 5.

3. **On linear algebras.** Let V be a vector space of n dimensions over a field F. Elements of V satisfy

$$\begin{pmatrix}\alpha_1\\ \cdot\\ \cdot\\ \cdot\\ \alpha_n\end{pmatrix} = (\alpha_i) = \sum \alpha_i d_i, \quad (\alpha_i) + (\beta_i) = (\alpha_i + \beta_i), \quad \alpha(\alpha_i) = (\alpha\alpha_i).$$

It is well known* that every F-operator endomorphism of V ($v \to v'$ implies $\alpha v \to \alpha v'$) is represented by an $n \times n$ matrix over F operating on V. For under such a map, $d_i \to \sum \alpha_{ji} d_j$, and

$$v = \sum \alpha_i d_i \to \sum \left(\sum \alpha_i \alpha_{ji}\right) d_j = Av,$$

where A is the matrix (α_{ij}). Now a linear associative algebra of order n over the field F is simply a ring $A(V)$ over V subject to the axioms (i) $\alpha(uv) = u(\alpha v)$ and (ii) $\alpha(uv) = (\alpha u)v$. Condition (i) requires that the endomorphism defined by the multiplier u be an F-operator map, that is, $uv = Uv$, where U is a matrix of the type just indicated. Hence in the correspondence of Theorem 1, $u \to U$; and by (ii), $\alpha u \to \alpha U$, $(\alpha \in F)$. Thus

$$A(V) \sim \Gamma \quad (\oplus, \odot; \Gamma, F)$$

where Γ is a subalgebra of the total $n \times n$ matrix algebra $\mathfrak{M}$ over F. This correspondence (which is the classical one) is biunique if and only if the left annulling ideal N of $A(V)$ is (0), a much weaker condition than the possession of unit element usually required. The Γ-operator property of the correspondence is significant in the light of the following remark, which is in part a result of Theorem 2:

If $V \sim \Gamma$ $(\oplus; \Gamma, F)$, Γ any subalgebra of $\mathfrak{M}$, then there exists an algebra $A(V)$ over V such that

$$A(V) \sim \Gamma \quad (\oplus, \odot; \Gamma, F).$$

That not every matrix representation of an algebra possesses the Γ-operator property is evinced by the example

$$A(V): \quad \begin{pmatrix}\alpha_1\\ \alpha_2\end{pmatrix}, \begin{pmatrix}\alpha_1\\ \alpha_2\end{pmatrix}\begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix} = \begin{pmatrix}\alpha_1 & \beta_1\\ \alpha_1 & \beta_2\end{pmatrix},$$

for

$$\begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix} \cong \begin{pmatrix}\beta_1 & \beta_2\\ 0 & 0\end{pmatrix} \quad (\oplus, \odot)$$

but the relation

* See van der Waerden, loc. cit., vol. 2, p. 111.

$$\begin{pmatrix}\alpha_1 & \alpha_2\\ 0 & 0\end{pmatrix}\begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix}\to\begin{pmatrix}\alpha_1 & \alpha_2\\ 0 & 0\end{pmatrix}\begin{pmatrix}\beta_1 & \beta_2\\ 0 & 0\end{pmatrix}$$

does not hold. However

$$\begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix}\sim\Gamma\equiv\begin{pmatrix}\beta_1 & 0\\ 0 & \beta_1\end{pmatrix}\ (\oplus,\odot;\Gamma).$$

4. **Reduction theorems for finite rings.** Let M be a module of order $m=p_1^{a_1}\cdots p_n^{a_n}$. Then $M=B_1+\cdots+B_n$ is a direct sum, B_i of order $p_i^{a_i}$, containing all elements of period dividing $p_i^{a_i}$. Moreover, $B_i=C_{i1}+\cdots+C_{ii_i}$, where C_{ij} is cyclic of order $p_i^{b_{ij}}$, $\sum_{j=1}^{i_i}b_{ij}=a_i$. The endomorphism ring Ω_M of M is a direct sum of endomorphism rings of the B_i:

$$\Omega_M=\Omega_1+\cdots+\Omega_n,$$

Ω_i a two-sided ideal in Ω_M, $\Omega_i\cap\Omega_j=\delta_{ij}\Omega_j$, $\Omega_i\Omega_j=\delta_{ij}\Omega_i^2$. Further, if $B=C_1+\cdots+C_t$, C_j of order p^{b_j}, be represented as a vector space

$$\begin{pmatrix}x_1\\ \cdot\\ \cdot\\ \cdot\\ x_t\end{pmatrix},\quad x_j\ (\mathrm{mod}\ p^{b_j}),\quad b_1\leqq\cdots\leqq b_t,$$

then Ω_B may be represented* by the ring of all matrices $(\beta_{jk})=(\alpha_{jk}p^{b_j-b_k})$, $p^{b_j-b_k}$ defined as 1 for $j<k$, β_{jk} reduced (mod p^{b_j}). Thus if M is represented as a vector space, Ω_M is a ring of matrices with blocks along the diagonal, the Ω_i-blocks having the (β_{jk}) structure described.†

THEOREM 3. *If* $M\sim\Gamma\subset\Omega_M\ (\oplus;\Gamma)$, *then* $\Gamma=\Gamma_1+\cdots+\Gamma_n$, *a direct sum of two-sided ideals in* Γ, *and*

$$B_i\sim\Gamma_i\subset\Omega_i\ (\oplus;\Gamma_i).$$

Let Γ_i be the map of B_i. Then Γ_i is a two-sided ideal in Γ, and every $\gamma\in\Gamma$ is a sum of $\gamma_i\in\Gamma_i$. Moreover $\Gamma_i\subset\Omega_i$. For let $b_i\to\lambda_i\in\Gamma_i$, $(\lambda_i=(\theta_1+\cdots+\theta_n),\ \theta_i\in\Omega_i)$. Since $b_i\in B_i$,

$$p_i^{a_i}b_i=0\to p_i^{a_i}(\theta_1+\cdots+\theta_n)=0.$$

Hence $p_i^{a_i}\theta_j=0$, $(j=1,\cdots,n)$. From the structure of Ω_i already indicated, $\theta_j=0$, $(j\neq i)$. Thus Γ is a direct sum.

* K. Shoda, *Über die Automorphismen einer endlichen Abelschen Gruppe*, Mathematische Annalen, vol. 100 (1928), p. 676.

† Note that B is admissible relative to Ω_M, that is, $\Omega_MB_i\subset B_i$.

THEOREM 4. *If* $M=B_1+\cdots+B_n$, $B_i\sim\Gamma_i$ $(\oplus;\Gamma_i)$, Γ_i *a subring of* Ω_i, *then* $\Gamma=\Gamma_1+\cdots+\Gamma_n$ *is direct,* Γ_i *a two-sided ideal in* Γ, *and*

$$M\sim\Gamma\subset\Omega_M\ (\oplus;\Gamma).$$

Since $\Gamma_i\subset\Omega_i$, Γ is a direct sum, and Γ_i is a two-sided ideal in Γ. Define $M\sim\Gamma$ by $m=b_1+\cdots+b_n\rightarrow\gamma_1+\cdots+\gamma_n$ (where $b_i\rightarrow\gamma_i$). Then addition is preserved. Let $\rho\in\Gamma$, $\rho=\mu_1+\cdots+\mu_n$, $(\mu_i\in\Gamma_i)$. Then

$$\begin{aligned}\rho m &= \rho b_1+\cdots+\rho b_n=\mu_1b_1+\cdots+\mu_nb_n\rightarrow\mu_1\gamma_1+\cdots+\mu_n\gamma_n\\ &=(\mu_1+\cdots+\mu_n)(\gamma_1+\cdots+\gamma_n).\end{aligned}$$

THEOREM 5. *Every ring over* $M=B_1+\cdots+B_n$ *is a direct sum of rings over the* B_i; *hence to construct all rings over* M *it is only necessary to construct all rings over the* B_i.

5. **On elementary modules.** M is said to be elementary in case there exists an isomorphism

$$M\cong\Omega_M\ (\oplus;\Omega_M).$$

THEOREM 6. M *is elementary if and only if there exists a ring with unit element,* $R(M)$ *over* M, *such that every endomorphism of* M *is defined by a left multiplier of* $R(M)$.

For if M is elementary, there exists a ring $R(M)$ such that

$$R(M)\cong\Omega_M\ (\oplus,\odot;\Omega_M)$$

where ab is defined as αb, $(a\longleftrightarrow\alpha)$. Let $m\rightarrow\theta m$ be an endomorphism of M. In the above isomorphism let $t\longleftrightarrow\theta$. Then $tm=\theta m$, $(t\in R(M))$. Conversely, if $R(M)$ is of this type,

$$R(M)\cong\Gamma\subset\Omega_M\ (\oplus,\odot;\Gamma),$$

and if one assumes $\theta\in\Omega_M$, there exists a $t\in R(M)$ such that $ta=\theta a$, $(a\in M)$. Hence $\theta\in\Gamma$ and $\Gamma=\Omega_M$; whence M is elementary.

COROLLARY. *The modules of rational numbers, and of rational integers* C (*the infinite cyclic group*) *are elementary.*

For it is readily shown that the only solution of the functional equation $\Phi=(a+b)=\Phi(a)+\Phi(b)$ in the field of rationals and the ring of integers is of the type $\Phi(a)=ra$ where r is a multiplier of the domain.

COROLLARY. *The only rings* $R(C)$ *over* C *are given by the multiplication* $a\cdot b$, *defined as any fixed positive integral multiple of the ordinary product* ab *in the ring of rational integers.*

To define a ring $R(C)$ we must obtain a homomorphism

$$C \sim \Gamma \ (\oplus; \Gamma)$$

where Γ is a subring of Ω_C, setting $a \cdot b = \alpha b$ $(a \to \alpha)$. But Ω_C is the ordinary ring of rational integers, its only subrings being principal ideals $\{m\}$. Hence we must have

$$C \sim \{m\} \ (\oplus; \{m\})$$

where $1 \to m$, $a \to ma$.

THEOREM 7. *If M is elementary, the units of Ω_M are in the centrum of Ω_M.**

For the endomorphism $\sigma^{-1}\Omega_M\sigma$ of the additive group of Ω_M (σ a unit) must be defined by a ring multiplier $\rho: \sigma^{-1}\Omega_M\sigma = \rho\Omega_M$. Then in particular $\sigma^{-1}\epsilon\sigma = \rho\epsilon$ and $\rho = \epsilon$.

COROLLARY. *A vector space V of order greater than or equal to 2 is not elementary.*

For there always exist nonsingular matrices not commutative with the total matrix algebra, and hence not in the centrum of Ω_V.

THEOREM 8. *A finite module M is elementary if and only if it is cyclic.*

For a cyclic M, Ω_M is represented by the $n \times n$ matrices $(\delta_{ij}\alpha_j)$, $\alpha_j \pmod{p_j^{e_j}}$. Hence under

$$\begin{pmatrix} \alpha_1 \\ \cdot \\ \cdot \\ \cdot \\ \alpha_n \end{pmatrix} \to \begin{pmatrix} \alpha_1 & & 0 \\ & \cdot & \\ & & \cdot \\ 0 & & \alpha_n \end{pmatrix},$$

M is elementary. If there are repeated primes in the type of M, then the order of Ω_M is greater than that of M and M is not elementary (see §4).

Thus the rings $R(M)$ over elementary finite M are completely known, $(\alpha_i)(\beta_i)$ being defined as $(\gamma_i\alpha_i\beta_i)$, $(0 \leq \gamma_i < p_i^{e_i})$.

UNIVERSITY OF WISCONSIN

* A stronger theorem holds: *If M is elementary, its endomorphism ring is commutative.*

CONCERNING MATRICES WITH ELEMENTS IN A COMMUTATIVE RING*

NEAL H. McCOY

1. **Introduction.** In a study of the algebraic properties of matrices with elements in a field, some of the most fundamental theorems are those having to do with the concepts of characteristic function and minimum function.† It is the purpose of the present note to suitably generalize some of the leading theorems in this connection to the case of matrices with elements in an arbitrary commutative ring R with unit element 1. For the most part, the theorems as well as the proofs are obtained by suitable generalizations of familiar theorems and proofs in the case in which the coefficient domain is restricted to be a field. However, the degree of generality here obtained seems to be of sufficient interest to warrant a brief account.

Let A denote a square matrix with elements in R. As will be indicated in §2, it is easy to define, in the usual way, the characteristic function $f(\lambda)$, and to show that $f(A)=0$. We shall call the principal ideal $(f(\lambda))$, in the ring $R[\lambda]$, the *characteristic ideal* of A.‡ The set of all elements $g(\lambda)$ of $R[\lambda]$ such that $g(A)=0$ is clearly an ideal in $R[\lambda]$ which may be called the *minimum ideal* of A. In general, this ideal will not be principal. The terms *characteristic ideal* and *minimum ideal* are used merely to emphasize the analogy with the characteristic and minimum functions of A in case the coefficient domain is a field. The actual determination of the minimum ideal is a fundamental problem, a solution of which is obtained in §3. The result, as stated in Theorem 3, is seen to be a generalization of the well known theorem of Frobenius concerning the minimum function. This theorem is the leading result of the present note.

One direction in which we propose to generalize certain familiar results will be sufficiently indicated by the remark that in place of irreducible factors of the characteristic (minimum) function of A we use the prime ideal divisors of the characteristic (minimum) ideal of A. For example, it is easy to show that the prime ideal divisors of the

* Presented to the Society, February 26, 1938.

† For known results concerning the characteristic and minimum functions, see J. H. M. Wedderburn, *Lectures on Matrices*, American Mathematical Society Colloquium Publications, vol. 17, 1934, chap. 2, or C. C. MacDuffee, *The Theory of Matrices*, chap 2. The former will be referred to hereafter as W, the latter as M.

‡ For fundamental definitions and properties of ideals, see van der Waerden, *Moderne Algebra*, or Krull, *Idealtheorie*.

minimum ideal coincide with those of the characteristic ideal. Some further miscellaneous results are given in §4.

2. **Some preliminary remarks.** Unless otherwise specifically stated, R will always denote an arbitrary commutative ring with unit element 1, and R_n will be used to denote the ring of all matrices of order n, with elements in R. Now R_n contains a subring isomorphic to R, which we shall identify with R, so that we shall not distinguish between the unit element of R_n and that of R. Throughout the paper, λ will be used as an indeterminate, and $R[\lambda]$ is the ring of polynomials in λ with coefficients in R. Similarly $R_n[\lambda]$ is the ring of polynomials in λ with coefficients in R_n or, from another point of view, the ring of matrices of order n with elements in $R[\lambda]$.

We now make an observation concerning algebraic identities.* Let C be the ring of rational integers, and consider the ring $C' = C[x_1, x_2, \cdots, x_m]$, where $x_1, x_2, \cdots, x_m$ are indeterminates. If now $f(x_1, x_2, \cdots, x_m)$ and $g(x_1, x_2, \cdots, x_m)$ are elements of C', the statement that in C'

$$f(x_1, x_2, \cdots, x_m) = g(x_1, x_2, \cdots, x_m),$$

implies that this is an algebraic identity, so that if $a_1, a_2, \cdots, a_m$ are any elements of R, then clearly

$$f(a_1, a_2, \cdots, a_m) = g(a_1, a_2, \cdots, a_m),$$

equality now being equality in R. It is understood that an integer k occurring as a coefficient in f or g is also to be replaced by $k \cdot 1$, where 1 is the unit element of R. As an application of these remarks, let $X = \|x_{ij}\|$ be a square matrix all of whose elements are indeterminates. Then in the ring $C[x_{11}, \cdots, x_{nn}]$, we have the familiar result that

$$(\text{adj } X)X = X(\text{adj } X) = |X|.$$

Now this equation is in reality a set of equations, each being of the type discussed above. Thus it follows that if D is a matrix with elements in any commutative ring with unit element, then

$$(\text{adj } D)D = D(\text{adj } D) = |D|. \tag{1}$$

It is obvious that many other theorems of this nature may be easily extended in like manner. We shall make use of some similar results without explicitly stating that they hold in the ring R.

It is now easy to prove the following theorem:

* Cf. K. Rychlik, *Eine Bemerkung zur Determinantentheorie*, Journal für die reine und angewandte Mathematik, vol. 167 (1932), p. 197.

THEOREM 1. *An element A of R_n has an inverse if and only if $|A|$ has an inverse in R.*

Suppose $|A|$ has an inverse b in R. Then from (1) it follows that $b(\text{adj } A)$ is the inverse of A in R_n. Conversely, if A has an inverse B, then from $AB=BA=1$, it follows by taking determinants that $|B|$ is an inverse of $|A|$ in R.

If A is an element of R_n, we call the polynomial

$$f(\lambda) = |\lambda - A| = \lambda^n + a_1\lambda^{n-1} + \cdots + a_n,$$

the characteristic function of A, and the principal ideal $(f(\lambda))$ in $R[\lambda]$ the *characteristic ideal* of A. It will be noted that $f(\lambda)$ is an element of $R[\lambda]$, the leading coefficient being 1 and the constant term being $\pm|A|$. Since the leading coefficient is 1, it is easy to see that, in $R[\lambda]$, $f(\lambda)$ is not a divisor of zero—a fact which will be used in §3.

Since, by (1), with $D=\lambda-A$, we have

$$(\lambda - A)\ \text{adj}\ (\lambda - A) = f(\lambda),$$

the factor theorem* shows that $f(A)=0$. We have thus proved the next theorem:

THEOREM 2. *If $f(\lambda)$ is the characteristic function of A, then $f(A)=0$.*

It is now easy to prove the following corollary:

COROLLARY.† *If A has an inverse in R_n, it is expressible as a polynomial in A with coefficients in R.*

3. **Determination of the minimum ideal.** As above, we let A denote a fixed element of R_n, the characteristic function of A being $f(\lambda)$. The *minimum ideal* of A is the ideal $\mathfrak{m}$ in $R[\lambda]$ of all polynomials $g(\lambda)$ such that $g(A)=0$.

Let the first minors of the matrix $\lambda-A$ be denoted by $h_{ij}(\lambda)$, $(i, j=1, 2, \cdots, n)$. We may now establish the following fundamental result:

THEOREM 3. *An element $g(\lambda)$ of $R[\lambda]$ is an element of $\mathfrak{m}$ if and only if*

$$g(\lambda)h_{ij}(\lambda) \equiv 0\ (f(\lambda)), \qquad i, j = 1, 2, \cdots, n. \tag{2}$$

Before proving the theorem, we may point out what it means in the case in which R is specialized to be a field. In that case, each ideal in $R[\lambda]$ is principal and thus $\mathfrak{m}=(m(\lambda))$, where $m(\lambda)$ is the minimum

* See A. A. Albert, *Modern Higher Algebra*, p. 26.

† The proof follows readily from Theorems 1 and 2 by the method of M, p. 21.

function of A. If $h(\lambda)$ is the g.c.d. of the $h_{ij}(\lambda)$, the theorem merely states that $m(\lambda)=f(\lambda)/h(\lambda)$, which is the theorem of Frobenius. We may remark also that in the language of ideal quotients,* Theorem 3 states that $\mathfrak{m}=(f(\lambda)):\mathfrak{h}$, where $\mathfrak{h}$ is the ideal with basis elements $h_{ij}(\lambda)$, $(i, j=1, 2, \cdots, n)$.

Our proof is a modification of a method of Perron.† Suppose $g(\lambda)$ satisfies all the conditions (2). Now the elements of adj $(\lambda-A)$ are precisely the $h_{ij}(\lambda)$ except possibly for sign. Thus we have

$$g(\lambda)\ \text{adj}\ (\lambda - A) = \|f(\lambda)k_{ij}(\lambda)\| = f(\lambda)K(\lambda), \tag{3}$$

where $K(\lambda)$ is a matrix with elements in $R[\lambda]$, in other words an element of $R_n[\lambda]$. If we multiply (3) by $\lambda-A$ on the left, and apply relation (1), we get

$$g(\lambda)f(\lambda) = f(\lambda)(\lambda - A)K(\lambda).$$

Now it was pointed out above that $f(\lambda)$ is not a divisor of zero, and we thus have $g(\lambda)=(\lambda-A)K(\lambda)$. The factor theorem then shows at once that $g(A)=0$, and thus that $g(\lambda)\equiv 0\ (\mathfrak{m})$.

Let us now assume that $g(\lambda)\equiv 0\ (\mathfrak{m})$. Then clearly

$$g(\lambda) = g(\lambda) - g(A) = (\lambda - A)G(\lambda),$$

where $G(\lambda)$ is an element of $R_n[\lambda]$. Multiplication by adj $(\lambda-A)$ yields the result

$$g(\lambda)\ \text{adj}\ (\lambda - A) = f(\lambda)G(\lambda),$$

and hence

$$g(\lambda)h_{ij}(\lambda) = \pm f(\lambda)g_{ij}(\lambda) \equiv 0\ (f(\lambda)), \qquad i, j = 1, 2, \cdots, n.$$

Thus the conditions (2) are satisfied, and the proof of the theorem is completed.

We may now prove the following corollary:

COROLLARY. *The prime ideal divisors of the minimum ideal of A coincide with those of the characteristic ideal of A.*

Since $f(\lambda)\equiv 0\ (\mathfrak{m})$, it is only necessary to prove that if $\mathfrak{p}$ is a prime ideal divisor of $(f(\lambda))$, it also divides $\mathfrak{m}$. Let $g(\lambda)$ be any element of $\mathfrak{m}$; we show that $g(\lambda)\equiv 0\ (\mathfrak{p})$. Taking determinants of both sides of relation (3), we get

* See W. Krull, *Ein neuer Beweis für die Hauptsätze der allgemeinen Idealtheorie*, Mathematische Annalen, vol. 90 (1923), pp. 55–64.

† See M, p. 20.

$$[g(\lambda)]^n[f(\lambda)]^{n-1} = [f(\lambda)]^n \mid K(\lambda) \mid,$$

from which it follows that

$$[g(\lambda)]^n = f(\lambda) \mid K(\lambda) \mid \equiv 0 \ (\mathfrak{p}).$$

But since $\mathfrak{p}$ is a prime ideal, this implies that $g(\lambda) \equiv 0 \ (\mathfrak{p})$.

4. **Further miscellaneous results.** We conclude with a few results which follow readily from the definition of the minimum ideal $\mathfrak{m}$.

THEOREM 4. *If $h(\lambda)$ is an element of $R[\lambda]$, then $h(A)$ has an inverse if and only if $(h(\lambda), \mathfrak{m}) = (1)$.*

The sufficiency of the condition is almost obvious. To prove the necessity of the condition, let us suppose that $h(A)$ has an inverse. Then, by the corollary to Theorem 2, it follows that there exists a polynomial $t(\lambda)$ in $R[\lambda]$ such that $t[h(A)]$ is the inverse of $h(A)$. That is, $h(\lambda)t[h(\lambda)] - 1 \equiv 0 \ (\mathfrak{m})$, and thus $(h(\lambda), \mathfrak{m}) = (1)$ as required.

Let $f'(\lambda)$ denote the formal derivative of the characteristic function $f(\lambda)$ of A. We now prove the following theorem:

THEOREM 5. *If $(f'(\lambda), \mathfrak{m}) = (1)$, the only matrices of R_n commutative with A are polynomials in A.*

Suppose B is an element of R_n commutative with A. An examination of the Sylvester identities* shows that they hold for the case in which the coefficient domain is our ring R. In particular, we have

$$f'(A)B = g(A),$$

where $g(A)$ is a polynomial in A with coefficients in R. Now since, by hypothesis, $(f'(\lambda), \mathfrak{m}) = (1)$, the preceding theorem shows that $f'(A)$ has an inverse, which is of necessity a polynomial in A. Thus B is of the required form.

THEOREM 6. *If $h(\lambda)$ is an element of $R[\lambda]$, then $h(A)$ is nilpotent if and only if $h(\lambda)$ is divisible by all prime ideal divisors of $\mathfrak{m}$.*

For clearly $h(A)$ is nilpotent if and only if some power of $h(\lambda)$ is in $\mathfrak{m}$, and a result of Krull† shows that this is the case if and only if $h(\lambda)$ is an element of each prime ideal divisor of $\mathfrak{m}$.

INSTITUTE FOR ADVANCED STUDY AND
SMITH COLLEGE

---

* W, p. 27.

† *Idealtheorie*, p. 9.

CREMONA TRANSFORMATIONS WITH AN INVARIANT RATIONAL SEXTIC

A. B. COBLE

It is well known that a Cremona transformation T with ten or fewer F-points will transform a rational sextic S into a rational sextic S' when the F-points of T are all located at the nodes of S. I have shown (cf. [1], p. 248, (9); [2], p. 255, (5)) that, even though the number of transformations T of the type indicated is infinite, the transforms S' are all included in $2^{13} \cdot 31 \cdot 51$ classes, the members of any one class being all projectively equivalent and a member of one class being projectively distinct from a member of another class. The sextic S itself is in one of these classes, T being then the identity. If S' is in the same class as S, and if C is the collineation which carries S' into S, then TC is a Cremona transformation of the same type as T which transforms S into itself. There is thus an infinite group of Cremona transformations which carry S into itself. If t is a parameter on S, the effect of an element of such a group on the points of S is represented by

$$t' = \frac{at + b}{ct + d}, \qquad ad - bc \neq 0. \tag{1}$$

It is an obvious question as to whether transformations T, other than the identical collineation, exist for which (1) reduces to $t' = t$; that is, whether S can be a locus of fixed points of a transformation T. I had expressed the *opinion* that such transformations T do not exist (cf. [1], end of §3). It was therefore most interesting to find in a recent article of G. Pompili (Pompili [1]) a purported construction of such a transformation. However the examination, made in the following, of this transformation shows that the construction is fallacious.

Let S be a generic rational sextic with nodes at $p_1, \cdots, p_7$ and at A, B, C. Let H be a generic member of the pencil (H) of elliptic sextics with nodes at $p_1, \cdots, p_7, B, C$, the pencil being determined by S and the square of the cubic $(p_1 \cdots p_7 BC)^2$. On H let g_B, g_C denote the pairs of points at the nodes. Then on this elliptic curve the equivalence

$$T\colon P' - P \equiv g_B - g_C \tag{2}$$

determines a birational correspondence which, extended over the various members of the pencil (H), yields a Cremona transformation T of the plane. If (H'), (H'') are similar pencils

$$(p_1^2 \cdots p_7^2 C^2 A^2)^6, \qquad (p_1^2 \cdots p_7^2 A^2 B^2)^6,$$

respectively, and if on H' and H'', respectively, the pairs of nodes C, A and A, B are represented by g_C', g_A' and g_A'', g_B'', then similar Cremona transformations T' and T'' exist which are determined by the equivalences

$$(3) \qquad T'\colon P' - P \equiv g_C' - g_A', \qquad T''\colon P' - P \equiv g_A'' - g_B''.$$

The three pencils (H), (H'), (H'') have one common member S.

Pompili remarks that on S we have the equivalences

$$(4) \qquad g_B \equiv g_A', \qquad g_C \equiv g_A'', \qquad g_C' \equiv g_B'',$$

whence it is "intuitive" that the transformation $R = T''T'T$ has S as a locus of fixed points. The fallacy seems to rest in the definition of T in (2). This definition is precise for every elliptic member of (H), but it becomes illusory for each of the twelve rational members of (H) including S. For, on a rational curve, we have $P' \equiv P$ and $g_B \equiv g_C$, and T in (2) takes the form $0 \equiv 0$. Even the conclusion that R has on S the expression $P' - P \equiv 0$ would give on S no information as to the relative position of P and P'.

The nature of the Cremona transformation R may be determined in the following way which also shows that S is not a locus of fixed points of R. Let b_A, b_B, b_C, be the three Bertini involutions with seven F-points at $p_1, \cdots, p_7$ and with their eighth F-points at A, B, C, respectively. By definition, b_B carries every 8-nodal sextic $(p_1^2 \cdots p_7^2 B^2)^6$ into itself and therefore carries every cubic of the pencil $(p_1 \cdots p_7 B)^3$ into itself. Thus the pairs of the involution b_B on each such 8-nodal sextic are cut out by the members of this pencil of cubics. In particular, H is such an 8-nodal sextic and g_C is one of the pairs of b_B. Thus the parametric expression of b_B on H with pairs PP'' is $P + P'' \equiv g_C$. Similarly the expression on H of b_C with pairs $P'P''$ is $P' + P'' \equiv g_B$. Hence $P' - P \equiv g_B - g_C$. Thus the transformation T in (2) which carries P into P' is $b_B b_C$. Hence the transformations T, T', T'', R used above are expressible in terms of Bertini involutions as follows:

$$(5) \qquad T = b_B b_C, \quad T' = b_C b_A, \quad T'' = b_A b_B, \quad R = b_A b_B \cdot b_C b_A \cdot b_B b_C.$$

It is easy to obtain the parametric equation on S of the transformation R. Let the nodal parameters of S at A, B, C, respectively, be the pairs $(at)^2$, $(bt)^2$, $(ct)^2$, and let the jacobians of these pairs be

$$(\alpha t)^2 = J(b, c), \qquad (\beta t)^2 = J(c, a), \qquad (\gamma t)^2 = J(a, b).$$

Let also r_α, r_β, r_γ be the binary involutions with fixed points $(\alpha t)^2$, $(\beta t)^2$, $(\gamma t)^2$, respectively. Then the Bertini involution b_A has B and C for fixed points and interchanges the nodal directions at each. Hence its effect on S is given by r_α. Thus the parametric equation on S of the transformation R is

$$R = r_\alpha r_\beta \cdot r_\gamma r_\alpha \cdot r_\beta r_\gamma . \tag{6}$$

Since S is generic, the quadratics $(at)^2$, $(bt)^2$, $(ct)^2$ are generic, and their three jacobians $(\alpha t)^2$, $(\beta t)^2$, $(\gamma t)^2$ are likewise three generic quadratics. Then $R \neq 1$; that is, S is not a locus of fixed points. For, if $R=1$, there exists a binary involution $r_\delta = r_\alpha r_\beta r_\gamma$ such that $r_\alpha r_\beta r_\gamma r_\delta = 1$. This would require that the jacobians of $(\alpha t)^2$, $(\beta t)^2$ and of $(\gamma t)^2$, $(\delta t)^2$ should coincide; that is, $(\alpha t)^2$, $(\beta t)^2$, $(\gamma t)^2$ would have a common apolar quadratic, and would therefore lie in a pencil and thus be non-generic.

Pompili ([1], p. 85) sets up the three transformations $r_\alpha r_\beta$, $r_\gamma r_\alpha$, $r_\beta r_\gamma$ and multiplies them together to get the result $R=1$. However, he uses the following order of multiplication, $r_\beta r_\gamma \cdot r_\gamma r_\alpha \cdot r_\alpha r_\beta$, which obviously yields the identity. *If this is his definition of* R, the corresponding Cremona transformation R is $b_B b_C \cdot b_C b_A \cdot b_A b_B$ which is, just as obviously, the ternary identical collineation rather than a Cremona transformation.

It may be observed that the Cremona transformation R defined in (5) has the order 2305. It has $p_1, \cdots, p_7$ as F-points of order 768 and A, B, C as F-points of respective orders 1008, 384, 144. The numerical matrix which describes its direct and inverse F-points and P-curves reads as follows:

		p_1	p_2	$\cdots$	p_7	A	B	C
	2305	−768	−768	$\cdots$	−768	−1008	−384	−144
p_1	768	−255	−256	$\cdots$	−256	−336	−128	−48
p_2	768	−256	−255	$\cdots$	−256	−336	−128	−48
.	.	.	.	. . .	.	.	.	.
p_7	768	−256	−256	$\cdots$	−256	−336	−128	−48
A	144	−48	−48	$\cdots$	−48	−63	−24	−8
B	384	−128	−128	$\cdots$	−128	−168	−63	−24
C	1008	−336	−336	$\cdots$	−336	−440	−168	−63

The first column of this table indicates that R transforms lines into curves of order 2305 with points of order 768 at $p_1, \cdots, p_7$ and with points of orders 144, 384, 1008 at A, B, C, respectively; the second column, that R transforms the set of directions at p_1 into a curve of order 768 with points of order 255 at p_1, of order 256 at

$p_2, \cdots, p_7$, and of orders 48, 128, and 336 at A, B, and C, respectively; and so on.

References

A. B. Coble

1. *The ten nodes of the rational sextic and the Cayley symmetroid*, American Journal of Mathematics, vol. 41 (1919), pp. 243–265.
2. *Algebraic Geometry and Theta Functions*, American Mathematical Society Colloquium Publications, vol. 10, New York, 1929.

Giuseppe Pompili

1. *Sulle trasformazioni cremoniane del piano che posseggono una curva di punti uniti*, Rendiconti del Seminario Matematico della Università di Roma, (4), vol. 2 (1938), pp. 47–87.

University of Illinois

A NOTE ON THE REDUCTION OF GENTZEN'S CALCULUS LJ*

H. B. CURRY

The reduction given by Gentzen† of his calculus LJ to the modified Heyting calculus LHJ is somewhat involved because he reduces everything to the axioms without assuming any knowledge of the calculus. By the use, however, of certain general theorems it is possible to simplify the reduction. The purpose of this note is to present an alternative reduction based on this principle. Although this new reduction may, if all the assumptions used are proved in detail from the axioms, conceivably be longer than Gentzen's, yet the formulas and principles established at the beginning (in §§1–4 below) are for the most part well known (or at least of some interest on their own account), and in terms of these the reduction (in §5) is almost immediate.

The new method has the further merit of showing, if we take the axioms of LHJ as a basis, that the schemes for implication follow from the axioms for implication only‡ and that those for conjunction, negation, and the quantifiers, respectively, involve only the axioms for implication and those for the operation concerned.§ It

* Presented at a joint meeting of the Society and the Association for Symbolic Logic, December 29, 1937.

† G. Gentzen, *Untersuchungen über das logische Schliessen*, Mathematische Zeitschrift, vol. 39 (1934), pp. 417–428.

‡ The scheme "Verdünnung im Sukzedens" is essentially a scheme for negation.

§ This statement requires that the formula 1.42 (below) be postulated as axiom of LHJ; to derive 1.42 from LHJ as it stands requires properties of conjunction.

then follows from Gentzen's principal theorem that if a formula can be established in the Heyting calculus at all, it can be established on the basis of those axioms of the calculus LHJ which concern implication and the other operations, if any, which actually appear in the formula.†

The notation of this paper is the same as Gentzen's except as follows: For conjunction $\wedge$ will be used instead of &; the usual notation (x) will be used for the universal quantifier; and ordinary (italic) letters, instead of German ones, will be used for unspecified expressions of the calculus. On the use of dots see my paper *On the use of dots as brackets in logical expressions.*‡ Certain formulas and theorems which are not necessary for the immediate purpose are preceded by *; some of these are useful for certain generalizations mentioned at the end.

1. **Preliminary formulas and principles.** The first step is to derive the following formulas and principles:

1.1. *Properties of implication.*

1.10. $\dfrac{A,\ A \supset B}{B}$.

*1.11. $A \supset A$.

1.12. $A \supset . B \supset A$.

1.13. $A \supset . A \supset B : \supset : A \supset B$.

1.14. $A \supset . B \supset C : \supset : B \supset . A \supset C$.

1.15. $B \supset C . \supset : A \supset B . \supset . A \supset C$.

1.16. $A \supset B . \supset : B \supset C . \supset . A \supset C$.

1.17. $\dfrac{A \supset B,\ B \supset C}{A \supset C}$.

1.2. *Properties of conjunction and alternation.* These are given, partly in dual form, as follows:

1.21. $A \wedge B . \supset . A,\ A \wedge B . \supset . B;\ B . \supset . A \vee B,\ A . \supset . A \vee B$.

1.22. $\dfrac{A,\ B}{A \wedge B}$.

1.23. $C \supset A . \wedge . C \supset B : \supset : C . \supset . A \wedge B$, and
$A \supset C . \wedge . B \supset C : \supset : A \vee B . \supset . C$.

† This result appears to be generally known; but I am acquainted with no published proof of it. For partial results of the same nature see D. Hilbert and P. Bernays, *Grundlagen der Mathematik*, vol 1, 1934, p. 71; also I. Johansson, Compositio Mathematica, vol. 4 (1937), p. 131. Added in proof: See also M. Wajsberg, *Untersuchungen über den Aussagenkalkul von A. Heyting*, Wiadomości Matematyczne, vol. 46 (1938), pp. 45–101.

‡ Journal of Symbolic Logic, vol. 2 (1937), pp. 26–28.

*1.24. $\dfrac{A \supset B}{A \wedge C \,.\supset.\, B \wedge C}$, $\dfrac{B \supset A}{B \vee C \,.\supset.\, A \vee C}$,

$\dfrac{A \supset B}{C \wedge A \,.\supset.\, C \wedge B}$; $\dfrac{B \supset A}{C \vee B \,.\supset.\, C \vee A}$.

*1.25. $A \supset . B \supset . A \wedge B$.

*1.26. $A \,.\wedge.\, A \supset B \,.\supset.\, B$.

*1.3. *Formal algebraic properties of* $\wedge$ *and* $\vee$, *that is, the commutative, associative, and distributive laws and the laws of tautology.*

1.4. *Properties of quantifiers.*

1.40. If Fa holds for a free variable a not in F, then $(x)Fx$ holds.

1.41. $(x)Fx \,.\supset.\, Fa$, and $Fa \,.\supset.\, (\exists x)Fx$.

1.42. If x does not occur in A,

$$(x) \,.\, A \supset Fx :\supset: A \,.\supset.\, (x)Fx, \quad (x) \,.\, Fx \supset A :\supset: (\exists x)Fx \,.\supset.\, A.$$

1.5. *Properties of negation.*

1.51. $A \supset . \neg A \supset \neg B$.†

1.52. $\dfrac{T}{A \supset \neg T \,.\supset.\, \neg A}$.

1.53. $\dfrac{T}{\neg T \,.\supset.\, A}$.

*1.54. $A \supset B \,.\supset.\, \neg B \supset \neg A$.

2. **A general substitution theorem.** Using these we may, following S. MacLane,‡ prove a general substitution theorem as follows:

2.1. First, let us say that an operation ϕ on A is positive when

$$\frac{A \supset B}{\phi A \supset \phi B}$$

is a valid scheme, negative when

$$\frac{A \supset B}{\phi B \supset \phi A}$$

† 1.51 is a weaker form of $A \supset . \neg A \supset B$, which is essentially equivalent to 1.53 (and, by 1.14, is equivalent to the axiom $\neg A \,.\supset.\, A \supset B$). However, 1.51 is true in the Minimalkalkul of Johansson (Compositio Mathematica, vol.4 (1936), pp. 119–136), while the stronger principles are not. Hence I shall use the stronger principle 1.53 only in the derivation of the scheme DS (5.43) where 1.51 does not suffice. This establishes all the schemes except DS for the calculus LM (Johansson, loc. cit., §5).

‡ *Abgekürzte Beweise im Logikkalkul*, Inaugural-dissertation, Göttingen 1934, p. 28 ff. A similar theorem appears in J. Herbrand, *Recherches sur la Théorie de la Démonstration*, Warsaw, 1930, §3.2.

is a valid scheme. Then by 1.15, 1.24 the operations $C\supset$,† $C\wedge$, $\wedge C$, $C\vee$, $\vee C$ are positive; on the other hand by 1.16, 1.54 $\supset C$ and $\neg$ are negative.‡

2.2. Suppose now we have an operation ψ which is compounded of certain operations $\phi_1, \phi_2, \cdots, \phi_n$ in the sense that

$$\psi A = \phi_n(\phi_{n-1}(\cdots(\phi_1 A)\cdots)).$$

Then if each of the ϕ_i is either positive or negative, ψ will be positive or negative according as the number of negative ϕ's is even or odd. (This is readily proved by induction.)

2.3. Hence suppose that we have a provable formula which is constructed from a particular instance of a formula A by the operations listed in 2.1; then if the number of negative operations is even, that particular instance of A may be replaced by any B such that $A\supset B$ holds; if the number of negative operations is odd, it may be replaced by any C such that $C\supset A$ holds.

3. **Further properties of ϕ.** Let us now consider two further properties of an operation ϕ as follows:

I. $\phi A\wedge\phi B\ .\supset.\ \phi(A\wedge B)$.

II. $(x)\phi(Fx)\ .\supset.\ \phi((x)Fx)$.

Concerning these we have the following theorems:

3.1. If the ϕ_i of 2.2 are all positive and have property I, then ϕ also has property I.

PROOF. Let ψ_k be defined by

$$\psi_1 A = \phi_1 A, \qquad \psi_{k+1}A = \phi_{k+1}(\psi_k A).$$

Then ψ_1 has property I. Suppose that ψ_k does; then by 1.1, and since ϕ_{k+1} has property I,

$$\begin{aligned}\psi_{k+1}A\wedge\psi_{k+1}B\ &.\supset.\ \phi_{k+1}(\psi_k A)\wedge\phi_{k+1}(\psi_k B)\\ &.\supset.\ \phi_{k+1}(\psi_k A\wedge\psi_k B)\\ &.\supset.\ \phi_{k+1}(\psi_k(A\wedge B))\end{aligned}$$

since ψ_k has property I and ϕ_{k+1} is positive. The transitive property of $\supset$ holds by 1.17. Hence ψ_{k+1} has property I, and the theorem follows by induction.

---

† By this I mean the operation converting A into $C\supset A$, and similarly for the other cases. The C's may be different in the different instances of the operators.

‡ The fact that $\neg$ is a negative operation is not used in establishing the schemes in §5 below. Hence 1.54 is marked with *; similarly for 1.24.

3.2. If the ϕ_i of 2.2 are all positive and have property II, then ψ also has property II. (The proof is similar to that of 3.1.)

3.3. In the following we shall need only the special case of the foregoing theorems where ϕ_i is $C_i \supset$, with the additional proviso, in the case of II, that C_i does not contain x. In this case the hypotheses regarding the ϕ_i are satisfied by 1.23 (in case I) and 1.42 (in case II).

4. **Chain implications.** Concerning formulas of the form

$$A_1 \supset . A_2 \supset . \cdots \supset . A_n \supset B \tag{1}$$

we have the following theorems:

4.1. From the axioms for implication (1.1) it follows that the A_i may be permuted in any manner, that repetitions may be cut out, and that new A's may be inserted at pleasure.

*4.2. From the axioms for implication and conjunction, it follows that (1) is equivalent to

$$A_1 \wedge A_2 \wedge \cdots \wedge A_n . \supset . B. \tag{2}$$

(The proof would require some of the starred propositions of §1.)

5. **Derivation of the schemes.** Suppose now that we replace the sequence

$$A_1, A_2, \cdots, A_n \rightarrow B \tag{3}$$

of Gentzen by the corresponding formula (1) of 4.1; with the proviso that where Gentzen has a sequence with a void consequent, a consequent Λ, where Λ is $\neg T$ and T is some provable formula, must first be supplied. Then the limitations of the calculus LJ insure that every sequence of an "LJ Herleitung" is of the kind just considered. We shall now see that Gentzen's schemes go over into valid principles of inference in the calculus LHJ. In this deduction I shall use D for "Verdünnung," Z for "Zusammenziehung," and T for "Vertauschung." Then the schemes are established as follows:

5.1. *Schemes for* $\supset$.

5.11. DA, ZA, TA, FES by 4.1.

5.12. Schnitt by 2.3.

5.13. FEA thus: by induction and 1.15, 1.17,

$$A \supset B . \supset : C_1 \supset . C_2 \supset . \cdots \supset . C_m \supset A . \supset : C_1 \supset . C_2 \supset . \cdots \supset . C_m \supset B.$$

Hence if $C_1 \supset . C_2 \supset . \cdots \supset . C_m \supset A$ holds, we have, by 1.14 and 1.10,

$$A \supset B . \supset . C_1 \supset . C_2 \supset . \cdots \supset . C_m \supset B.$$

The scheme is now reduced to Schnitt.

5.2. *Schemes for* $\wedge$ *and* $\vee$.
5.21. UES by 1.23, 3.1, 3.3.
5.22. OEA by 1.23 immediately.
5.23. UEA and OES by 1.21 and 2.3.
5.3. *Schemes for quantifiers.*
5.31. AES by 1.40, 3.2, 3.3.
5.32. EEA by 1.40, 1.42, 1.10.
5.33. AEA and EES (dually) by 1.41 and 2.3.
5.4. *Schemes for negation.*
5.41. NEA thus:

$$\frac{C_1 \supset . C_2 \supset . \cdots \supset . C_m \supset A}{C_1 \supset . C_2 \supset . \cdots \supset . C_m \supset . \neg A \supset \Lambda} \text{ (1.51, 2.3)}$$

$$\frac{C_1 \supset . C_2 \supset . \cdots \supset . C_m \supset . \neg A \supset \Lambda}{\neg A \supset . C_1 \supset . C_2 \supset . \cdots \supset . C_m \supset . \Lambda.} \text{ (4.1)}$$

5.42. NES thus:

$$\frac{A \supset . C_1 \supset . C_2 \supset . \cdots \supset . C_m \supset \neg T}{C_1 \supset . C_2 \supset . \cdots \supset . C_m \supset . A \supset \neg T} \text{ (4.1)}$$

$$\frac{C_1 \supset . C_2 \supset . \cdots \supset . C_m \supset . A \supset \neg T}{C_1 \supset . C_2 \supset . \cdots . C_m \supset . \neg A.} \text{ (1.52, 2.3)}$$

5.43. DS by 1.53, 2.3.

6. **Conclusion.** It may be worth remarking that by these same methods many of Gentzen's schemes may be established without the restriction to not more than one element in the consequent; the consequent must then be taken as a logical sum. In fact by the use of the formulas marked with a * this can be done for all the schemes which do not involve quantifiers except NES and FES.† These schemes are in general invalid. NES, in fact, leads to the law of excluded middle (Gentzen, loc. cit., p. 193); while FES leads (for void Γ) to the scheme

$$\frac{A . \supset . C \vee B}{C . \supset . A \vee B},$$

which is not valid.

THE PENNSYLVANIA STATE COLLEGE

† In the minimal calculus the generalized NEA is also invalid.

SOME INVARIANTS UNDER MONOTONE TRANSFORMATIONS*

D. W. HALL† AND A. D. WALLACE

We assume that S is a locally connected, connected, compact metric space and that P is a property of point sets. For any two points a and b of S we denote by $C(ab)$ (respectively $C_i(ab)$) a closed (closed irreducible) cutting of S between the points a and b. We consider the following properties:

$\Delta_0(P)$. *If S is the sum of two continua, their product has property P.*

$\Delta_1(P)$. *If K is a subcontinuum of S and R is a component of $S-K$, then the boundary of R, $(F(R)=\overline{R}-R)$, has property P.*

$\Delta_2(P)$. *Each $C_i(ab)$ has property P.*

$\Delta_3(P)$. *If A and B are disjoint closed sets containing the points a and b, respectively, there is a $C(ab)$ disjoint from $A+B$ and having property P.*

If P is the property of being connected, the four properties $\Delta_i(P)$ are equivalent as shown by Kuratowski.‡ Indeed it may be seen that Kuratowski's proofs allow us to state the following theorem:

THEOREM 1. *For any property P of point sets, $\Delta_i(P)$ implies $\Delta_{i+1}(P)$ for $i=0, 1, 2$.*

This result is the best possible in the sense that there is a property (that of being totally disconnected) for which no other implication holds.

The single-valued continuous transformation $T(S)=S'$ is said to be *monotone* if the inverse of every point is connected. It may be seen that the following statements are true:§

(i) *The inverse of every connected set is connected.*

(ii) *If the set X separates S between the inverses of the points x and y, then $T(X)$ separates S' between x and y.*

THEOREM 2. *If the property P is invariant under monotone trans-*

---

* Presented to the Society, October 29, 1938.

† National Research Fellow.

‡ C. Kuratowski, *Une caractérisation topologique de la surface de la sphère*, Fundamenta Mathematicae, vol. 13 (1929), p. 307, and references given there.

§ G. T. Whyburn, *Non-alternating transformations*, American Journal of Mathematics, vol. 56 (1934), p. 294.

formations, then for each $i=0, 1, 2, 3$, *the property* $\Delta_i(P)$ *is invariant under the monotone transformation* $T(S)=S'$.

PROOF. (0) If $S'=L+M$, the summands being continua, then $S=L^{-1}+M^{-1}$* is a sum of continua. Hence the set $L^{-1}\cdot M^{-1}$ has property P and $L\cdot M=T(L^{-1}\cdot M^{-1})$ then has property P.

(1) If R is a component of $S'-K$, where K is a continuum, then R^{-1} is a component of the complement of the continuum K^{-1}. By assumption, $F(R^{-1})$ has property P. It follows that its image has property P. But we have $T(F(R^{-1}))=T(\overline{R^{-1}}-R^{-1})=\overline{T(R^{-1})}-R=F(R)$.

(2) Assume that C is a $C_i(ab)$ in S'. From the continuity of T it follows that C^{-1} is a $C(pq)$ in S, p and q being any two points in the inverses of a and b, respectively. Since the inverses of a and b are connected, there exists a cutting K of S† between these two sets such that K is a $C_i(xy)$, where $T(x)=a$ and $T(y)=b$; and further K is a subset of C^{-1}. Thus K has property P; hence $T(K)$ has. But $T(K)\subset C$, and $T(K)$ is a $C(ab)$. It follows that $T(K)=C$ and from this that C has property P.

(3) Let A and B denote disjoint closed subsets of S' containing a and b. If x and y are points which map into a and b, then by hypothesis there is a cutting K of S between x and y that is disjoint with A^{-1} and B^{-1} and has property P. Since, clearly, K is a cutting of S between the inverses of a and b, it follows that $T(K)$ cuts S' between a and b, is disjoint with $A+B$, and has property P.

As an application we have the following known results:‡

THEOREM 3. *The property of a locally connected continuum to be a dendrite, a regular curve, or a rational curve is a monotone invariant.*

To see this we take P to be the property of being a point, a finite set of points, or a countable set of points and apply the invariance of $\Delta_3(P)$.

THE UNIVERSITY OF VIRGINIA

---

* If X is a subset of S', we denote by X^{-1} the inverse of X.

† G. T. Whyburn, *Concerning irreducible cuttings of continua*, Fundamenta Mathematicae, vol. 13 (1929), p. 42.

‡ See the fourth footnote and references given there.

REGULARITY OF FUNCTION-TO-FUNCTION TRANSFORMATIONS*

MAHLON M. DAY

1. **Introduction.** In a recent note Hill† considered the transformation

$$U_x(f) = \int_0^\eta K(x, y)f(y)dy \tag{1}$$

operating on the class $\mathfrak{B}_1$ of measurable, essentially bounded, complex-valued functions f of one real variable y defined for $0<y<\eta$, and satisfying the condition that $\lim_{y\to\eta} f(y)$ exists. The kernel $K(x, y)$ is defined for $0<x<\xi$, $0<y<\eta$, and the integral is interpreted in the Lebesgue sense. Hill derived a set of conditions on $K(x, y)$ necessary and sufficient that the transformation (1) be *regular* on $\mathfrak{B}_1$, that is, that for every f in $\mathfrak{B}_1$, $U_x(f)$ be defined for all x, and $\lim_{x\to\xi} U_x(f) = \lim_{y\to\eta} f(y)$.

In §2 of the present paper we generalize Hill's results for a transformation on a class $\mathfrak{B}_m$ of bounded measurable functions of m real variables to a class of functions of n real variables. This transformation can be expressed in the form (1) with x standing for $x^1, x^2, \cdots, x^n$ and y for $y^1, y^2, \cdots, y^m$. In §3 we define for each kernel $K(x, y)$ its domain of regularity $\mathfrak{R}$ as the largest class of functions on which (1) is regular, and we determine some conditions necessary and sufficient that a function f be in $\mathfrak{R}$. We employ these results in §4 to derive conditions on $K(x, y)$ necessary and sufficient for the transformation (1) to be regular on certain classes of functions more inclusive than $\mathfrak{B}_m$. Finally, in §5, we consider several particular cases of the problem of determining a kernel with a specified class of functions as its domain of regularity.

2. **Hill's theorem in many variables.** We use the following notation throughout this paper: x stands for the ordered set of n real variables $x^1, x^2, \cdots, x^n$, and y for $y^1, y^2, \cdots, y^m$; for 0, 0, $\cdots$, 0 we write 0. The equality $a=b$ means that for each j, $a^j=b^j$; $a<b$ that for each j, $a^j<b^j$; $a\not<b$ that for at least one j, $a^j\geqq b^j$; and $a>b$ that $b<a$. We define the *interval* (a, b) as the set of points c such that $a<c<b$.

---

* Presented to the Society, September 6, 1938.

† J. D. Hill, *A theorem in the theory of summability*, this Bulletin, vol. 42 (1936), pp. 225–228.

We consider only finite-valued functions defined on a given interval $0<y<\eta$ or $0<x<\xi$ where we allow ξ and η to be chosen so that any of the ξ^i or η^j may be either positive real numbers or may be infinite; in either case if $y<\eta$, then every y^j is finite.

We let $\lim_{y\to\eta} f(y)$ stand for the multiple limit, if it exists,* of $f(y)=f(y^1, \cdots, y^m)$ as the y^j tend to the η^j simultaneously but independently. We interpret all integrals as m-dimensional Lebesgue integrals which, by Fubini's theorem,† can be evaluated, whenever they exist, by iterated integration.

DEFINITION 1. *On a given interval* $(0, \eta)$ *let* $\mathfrak{W}_m$ *be the class of complex-valued functions,* $f(y)=f_1(y)+if_2(y)$, *satisfying the conditions that* (a) f_1 *and* f_2 *are measurable on* $(0, \eta)$, *and* (b) $f(y)$ *tends to a finite limit* L_f *as* y *tends to* η.

Let $\mathfrak{B}_m$ *be the subclass of* $\mathfrak{W}_m$ *such that* $f \in \mathfrak{B}_m$ *if and only if* f *is essentially bounded on* $(0, \eta)$.

DEFINITION 2. *For a given* ξ *and* η *let* $K(x, y)=K_1(x, y)+iK_2(x, y)$ *be a complex-valued function defined for* $0<x<\xi$ *and* $0<y<\eta$ *and such that for each* x, $K(x, y)$ *is a measurable function of* y. *Let*

$$
\begin{aligned}
U_x(f) &= \int_0^\eta K(x, y)f(y)dy \\
&= \int_0^\eta [K_1f_1 - K_2f_2] + i\int_0^\eta [K_1f_2 + K_2f_1].
\end{aligned} \tag{1}
$$

Let $\mathfrak{W}_m'$ *be any subclass of* $\mathfrak{W}_m$; $K(x, y)$ *is said to define a regular transformation on* $\mathfrak{W}_m'$, *or is said to be regular on* $\mathfrak{W}_m'$, *if and only if for each* f *in* $\mathfrak{W}_m'$, (a) *for each* x *in* $(0, \xi)$, $U_x(f)$ *exists, and* (b) $\lim_{x\to\xi} U_x(f)=L_f$.

THEOREM 1. *For* $K(x, y)$ *to be regular on* $\mathfrak{B}_m$, *it is necessary and sufficient that* (a) $\int_0^\eta K(x, y)dy\to 1$ *as* $x\to\xi$, (b) *for every measurable set* $H\subset(0, \eta)$ *not having* η *as a limit point* $\int_H K(x, y)dy\to 0$ *as* $x\to\xi$, (c) *for each* x *in* $(0, \xi)$, $\int_0^\eta |K(x, y)|dy$ *exists, and* (d) *there exist a number* M *and a point* $X<\xi$ *such that* $\int_0^\eta |K(x, y)|dy<M$ *for all* x *in* (X, ξ).

The proof of this theorem will not be given here since the notation which makes it possible to state the theorem in this form also suggests the necessary minor modifications in Hill's proof of the special case in which $n=m=1$.

---

* "Exists" shall mean "is a finite real or complex number."

† See S. Saks, *Theory of the Integral*, p. 77.

3. **The domain of regularity of a transformation.** We state the following definitions:

DEFINITION 3. *For a given kernel $K(x, y)$ the domain of regularity $\mathfrak{R}$ of* (1) *is the class of all functions f which satisfy the conditions*:
(2) $f \in \mathfrak{B}_m$.
(3) *For every x in $(0, \xi)$, $U_x(f)$ exists.*
(4) $\lim_{x\to\xi} U_x(f) = L_f$.

DEFINITION 4. *For any function $f(y)$ and any real number b let $B = E_y[|f(y)| > b]$, and let $f_b(y) = 0$ if $y \in B$, $f_b(y) = f(y)$ otherwise.*

We remark that if $f(y)$ is in $\mathfrak{B}_m$, and if $b > |L_f|$, then $f_b(y)$ is in $\mathfrak{B}_m$ and $L_f = L_{f_b}$.

THEOREM 2. *If $K(x, y)$ is regular on $\mathfrak{B}_m$, a function f belongs to $\mathfrak{R}$ if and only if it satisfies the conditions* (2) *and* (3), *and the following condition*:
(5) *For every $b > |L_f|$,*

$$\int_B K(x, y)f(y)dy \to 0 \quad as \quad x \to \xi.$$

PROOF. We need only show (4) and (5) equivalent here. We have for any $b > |L_f|$,

$$U_x(f) = U_x(f_b) + \int_B K(x, y)f(y)dy.$$

But $f_b \in \mathfrak{B}_m$ and $L_{f_b} = L_f$; so $\lim_{x\to\xi} U_x(f_b) = L_f$, and in order that $U_x(f) \to L_f$ as $x \to \xi$ it is necessary and sufficient that $\int_B K(x, y)f(y)dy \to 0$ as $x \to \xi$.

THEOREM 3. *If $K(x, y)$ is regular on $\mathfrak{B}_m$, a function f belongs to $\mathfrak{R}$ if and only if it satisfies the conditions* (2) *and* (3), *and the additional condition*:
(6) *For every $\epsilon > 0$, there exist a number $b > |L_f|$ and a point $X < \xi$ such that for all x in (X, ξ)*

$$\left| \int_B K(x, y)f(y)dy \right| < \epsilon.$$

PROOF. Here we need only show (4) and (6) equivalent. If K is regular on $\mathfrak{B}_m$ and f satisfies (4), by the preceding theorem, f satisfies (5) and so obviously (6).

For any $b > |L_f|$,

$$| U_x(f) - L_f | \leqq \left| \int_B Kf \right| + | U_x(f_b) - L_f |.$$

If f satisfies (6) and is regular on $\mathfrak{B}_m$, then for any $\epsilon > 0$, there exist a $b > |L_f|$ and an $X < \xi$ such that the right-hand side is less than ϵ for x in (X, ξ). But the left-hand side does not depend on b; so $U_x(f) \to L_f$ as $x \to \xi$ and (6) implies (4).

4. Conditions for regularity on other subclasses of $\mathfrak{B}_m$. In this section we consider conditions on $K(x, y)$ necessary and sufficient that the transformation (1) be regular on certain classes containing $\mathfrak{B}_m$.

DEFINITION 5. *Let $\mathfrak{S}_m$ be the subclass of $\mathfrak{B}_m$ such that $f \in \mathfrak{S}_m$ if and only if for every $Y < \eta$, $f(y)$ is L-integrable over $(0, Y)$. Let $\mathfrak{T}_m$ be the subclass of $\mathfrak{S}_m$ such that $f \in \mathfrak{T}_m$ if and only if there exists a point $Y < \eta$ such that f is essentially bounded in the region $(0, \eta) - (0, Y)$. Let $\mathfrak{R}_m$ be the subclass of $\mathfrak{B}_m$ such that $f \in \mathfrak{R}_m$ if and only if for every $Y < \eta$, $f(y)$ is essentially bounded on $(0, Y)$.*

We have the obvious relations that $\mathfrak{B}_m \supset \mathfrak{S}_m \supset \mathfrak{T}_m \supset \mathfrak{B}_m$, that $\mathfrak{S}_m \supset \mathfrak{R}_m \supset \mathfrak{B}_m$, that $\mathfrak{R}_1 = \mathfrak{B}_1$, and that $\mathfrak{T}_1 = \mathfrak{S}_1$.

THEOREM 4. *For $K(x, y)$ to be regular on $\mathfrak{T}_m$ it is necessary and sufficient that $K(x, y)$ be regular on $\mathfrak{B}_m$, and satisfy the following condition:*

(7) *For every $Y < \eta$,* (a) *for each x,*

$$F_Y(x) = \operatorname*{ess\,sup}_{0 < y < Y} | K(x, y) |$$

exists, and (b) *there exists a number M and a point $X < \xi$ such that $F_Y(x) < M$ for every x in (X, ξ).*

PROOF OF NECESSITY. The first condition is necessary as $\mathfrak{B}_m$ is contained in $\mathfrak{T}_m$. If (7a) is not satisfied, there exist $Y < \eta$ and $x < \xi$ such that $K(x, y)$ is not essentially bounded on $(0, Y)$. Hence there exists a function $f(y)$, L-integrable in $(0, Y)$ and zero in $(0, \eta) - (0, Y)$. such that

$$U_x(f) = \int_0^Y K(x, y) f(y) dy$$

does not exist. But this function is in $\mathfrak{T}_m$; so (7a) is necessary.

For the necessity of (7b) we again modify Hill's proof. Fix any $Y < \eta$, and let $\mathfrak{L}_m$ be the subclass of $\mathfrak{T}_m$ such that f is in $\mathfrak{L}_m$ if and only if $f(y) = 0$ for y in $(0, \eta) - (0, Y)$. Then with the usual definition of

equality, addition, and multiplication by real numbers, $\mathfrak{L}_m$ is a Banach space with norm given by $\|f\| = \int_0^Y |f(y)|\,dy$. For each fixed x,

$$U_x(f) = \int_0^Y K(x, y)f(y)dy$$

is a linear functional on $\mathfrak{L}_m$ with norm $\|U_x\| = F_Y(x)$.

For any sequence $\{x_k\}$ tending to ξ, let $U_{x_k} = U_k$. Then for every f in $\mathfrak{L}_m$, $U_k(f) \to 0$ as $k \to \infty$ since $K(x, y)$ is regular on $\mathfrak{T}_m$. Hence, for every f in $\mathfrak{L}_m$, the sequence $\{|U_k(f)|\}$ is bounded; so, by a theorem of Banach,* the sequence of norms $\{\|U_k\|\}$ is bounded. Since $\{x_k\}$ is any sequence tending to ξ, and since $\|U_x\| = F_Y(x)$, there exist M and $X < \xi$ such that $F_Y(x) < M$ for x in (X, ξ).

PROOF OF SUFFICIENCY. For any f in $\mathfrak{T}_m$, there exist M and $Y < \eta$ such that $|f(y)| < M$ for y in $(0, \eta) - (0, Y)$. Then for any $b > M$ the set B lies in $(0, Y)$ and

$$\int_0^\eta |K(x, y)f(y)|\,dy \leqq \int_0^\eta |Kf_b| + \int_B |Kf|$$

$$\leqq b\int_0^\eta |K| + F_Y(x)\int_B |f|.$$

Both integrals in the last term exist; so $\int_0^\eta |Kf|$ exists and $U_x(f)$ therefore exists for each x. Now $|\int_B Kf| \leqq F_Y(x)\int_B |f|$, and this can be made arbitrarily small by taking x in (X, ξ) and b sufficiently large. By Theorem 3, f is in $\mathfrak{K}$; therefore we have $\mathfrak{T}_m \subset \mathfrak{K}$, and $K(x, y)$ is regular on $\mathfrak{T}_m$.

DEFINITION 6. *In $(0, \eta)$ a set E is said to be essentially contained in a set F if and only if $m(E - EF) = 0$.*

DEFINITION 7. *For any given kernel $K(x, y)$ for each $x < \xi$ and each $Y < \eta$ let Y_x be the point such that* (a) *the sum of the intervals $(0, Y_x)$ and (Y, η) essentially contains $E_y[K(x, y) \neq 0]$, and* (b) *if $(0, Y_0) + (Y, \eta)$ essentially contains $E_y[K(x, y) \neq 0]$, then $(0, Y_x) \subset (0, Y_0)$.*

THEOREM 5. *For $K(x, y)$ to be regular on $\mathfrak{S}_m$ it is necessary and sufficient that $K(x, y)$ be regular on $\mathfrak{T}_m$ and satisfy the condition:*

(8) *For every $Y < \eta$,* (a) *for every x in $(0, \xi)$, $Y_x < \eta$, and* (b) *there exist $Y' < \eta$ and $X < \xi$ such that $Y_x < Y'$ for all x in (X, ξ).*

PROOF OF NECESSITY. If $m = 1$ condition (8) is satisfied by every kernel $K(x, y)$, since, for each Y and x, there is always a $Y_x \leqq Y$ such

* S. Banach, *Théorie des Opérations Linéaires*, p. 80, Theorem 5.

that $(0, Y_x)+(Y, \eta)$ essentially contains $E_y[K(x, y)\neq 0]$. (8b) can then be satisfied by taking $X=0$ and $Y'\geqq Y$.

For $m>1$ fix any $Y<\eta$; if (8a) is not satisfied for some x, then there exist a sequence $\{Y_k\}\to\eta$ and a function f_0 in $\mathfrak{S}_m$ such that $\int_0^{Y_k} Kf_0 > k$. Then $U_x(f_0)$ does not exist; hence (8a) is necessary.

If (8b) is not satisfied, there is a $Y<\eta$ such that for each $Y'<\eta$ and $X<\xi$ there is an x in (X, ξ) such that $Y_x \not< Y'$. Hence there exists a sequence $\{x_k\}\to\xi$ such that, if we let $Y_{x_k}=Y_k$, $(0, Y_k)-(0, Y_{k-1})$ is not empty. Moreover $\{x_k\}$ can be so chosen that for some $j\leqq m$, we have $Y_k^j\to\eta^j$ and $Y_k^j > Y_{k-1}^j$. Then there exists a function f in $\mathfrak{S}_m$ such that $f(y)=0$ for y in (Y, η) and such that for each k,

$$U_k(f) = \int_0^{Y_k} K(x_k, y)f(y)dy > 1.$$

This is possible because for each k there is a part of $(0, Y_k)$ which is in none of the sets $(0, Y_p)+(Y, \eta)$ for any $p<k$, and in this part of $(0, Y_k)$ there is a set of positive measure on which $K(x_k, y)\neq 0$. Then $L_f=0$, but $U_k(f)>1$ for all k so $U_x(f)$ cannot tend to 0 as $x\to\xi$. This contradiction proves (8b) necessary.

PROOF OF SUFFICIENCY. Condition (8a) and the condition that K be regular on $\mathfrak{T}_m$ insure that $U_x(f)$ exists for all f in $\mathfrak{S}_m$ and all x in $(0, \xi)$. For any f in $\mathfrak{S}_m$, there exists $Y<\eta$ and M such that $|f(y)|<M$ for y in (Y, η), because $f(y)\to L_f$ as $y\to\eta$. Then by (8b) there exist $Y'<\eta$ and $X<\xi$ such that for each x in (X, ξ), $K(x, y)=0$ almost everywhere outside $(0, Y')+(Y, \eta)$. Let $f_0(y)=f(y)$ for y in $(0, Y')+(Y, \eta)$ and $f_0(y)=0$ elsewhere. Then f_0 is in $\mathfrak{T}_m$ since it is in $\mathfrak{S}_m$ and $|f_0(y)|<M$ for y in $(0, \eta)-(0, Y')$. Then $L_{f_0}=L_f$; so $U_x(f_0)\to L_f$ as $x\to\xi$. But for x in (X, ξ), $U_x(f)=U_x(f_0)$; so $\lim_{x\to\xi} U_x(f)=L_f$ and $K(x, y)$ is regular on $\mathfrak{S}_m$.

THEOREM 6. *For $K(x, y)$ to be regular on $\mathfrak{R}_m$ it is necessary and sufficient that $K(x, y)$ be regular on $\mathfrak{B}_m$ and satisfy condition* (8).

The proof is almost the same as that of Theorem 5.

We remarked that the classes $\mathfrak{T}_1$ and $\mathfrak{S}_1$ are identical. It follows that (8) is not a restriction on $K(x, y)$ for $m=1$. From Theorems 4 and 5 we see that for $m>1$, the generalized arithmetic mean transformation given by $K(x, y)=1/x^1x^2\cdots x^m$ if $0<y^j<x^j$ for all $j\leqq m$, and $K(x, y)=0$ otherwise, is regular on $\mathfrak{T}_m$ but not on $\mathfrak{S}_m$ or $\mathfrak{R}_m$. This result resembles that of Kojima* on summability of multiple sequences.

* Kojima, *On the theory of double sequences*, Tôhoku Mathematical Journal, vol. 21 (1922), pp. 3–14.

5. **Kernels with specified domains of regularity.** In this section we consider the problem of defining a kernel having as domain of regularity one of the classes $\mathfrak{W}_m$, $\mathfrak{S}_m$, and $\mathfrak{B}_m$. For convenience we shall consider $n=m$ and all η^j and ξ^j infinite, but analogous results hold for any choice of these limits.

THEOREM 7. *For no m is $\mathfrak{W}_m$ the domain of regularity of any kernel.*

PROOF. If for each f in $\mathfrak{W}_m$ and each x in $(0, \xi)$, $U_x(f)$ exists, then for each x, $K(x, y)=0$ for almost all y in $(0, \eta)$. Hence $\int_0^\eta K(x, y)dy=0$, so

$$\lim_{x\to\xi}\int_0^\eta K(x, y)dy = 0;$$

but $\mathfrak{W}_m \supset \mathfrak{B}_m$, so by Theorem 1,

$$\lim_{x\to\xi}\int_0^\eta K(x, y)dy = 1.$$

THEOREM 8. *The kernel*

$$K(x, y)=\begin{cases}\prod_{j=1}^m(1-x^j) \text{ if, for every } j,\ 0<x^j<1 \text{ and } 0<y^j<1/(1-x^j),\\ 2^m\prod_{j=1}^m 1/x^j \text{ if, for some } j,\ 1\leqq x^j<\infty \text{ and for every } j,\\ \quad x^j/2<y^j<x^j,\\ 0 \text{ otherwise,}\end{cases}$$

has $\mathfrak{S}_m$ for domain of regularity.

PROOF. For each x, $\int_0^\eta K(x, y)dy=1$, and for all x and y, $0\leqq K(x, y)\leqq 1$. From this it is easily verified that K satisfies the conditions of Theorem 5 so that $\mathfrak{S}_m \subset \mathfrak{K}$. If f is not in $\mathfrak{S}_m$, then there exists a $Y<\eta$ such that $\int_0^Y f$ does not exist; also there is an x_0 such that for each j, $0<x_0{}^j<1$ and $1/(1-x_0{}^j)\geqq Y^j$ so that

$$\int_0^Y K(x_0, y)f(y)dy$$

does not exist and f is not in $\mathfrak{K}$. Hence $\mathfrak{K}=\mathfrak{S}_m$.

THEOREM 9. *There is a kernel having $\mathfrak{B}_m$ as its domain of regularity.*

PROOF. Let $\{G\}$ be the class of open sets in $(0, \eta)$; then there is a one-to-one correspondence between the class of points of the interval $0<s<1$ and the class of sequences $\{G_i\}$ of sets of $\{G\}$. Let $g_i(y)$ be the characteristic function of G_i. For each s and corresponding sequence $\{G_i\}$ let $f_s(y)=\sum_{i=0}^\infty g_i(y)$; then define

$$K(x, y) = \begin{cases} f_s(y) \text{ if } x^1 = s < 1 \text{ and } \int_0^\infty f_s < \infty, \\ \prod_{j=1}^m 1/x^j \text{ if } x^1 \geqq 1 \text{ and } 0 < y^j < x^j \text{ for all } j, \\ 0 \text{ otherwise.} \end{cases}$$

Then K is regular on $\mathfrak{B}_m$; so $\mathfrak{K} \supset \mathfrak{B}_m$. If $f \in \mathfrak{W}_m - \mathfrak{B}_m$, either there is a $b > 0$ such that B is of finite measure and $\int_B f$ exists or not; if there is no such b, $|f|$ dominates a function with this property. Hence there is a function $f_0(y)$, zero if y is not in B, such that $\int_B f_0$ exists and $\int_B f_0 f$ does not. Then there is an $s < 1$ such that f_s dominates f_0 but $\int_0^\infty f_s < \infty$. Hence there is an x such that $U_x(f)$ does not exist; so $\mathfrak{B}_m \supset \mathfrak{K}$.

For $m \geqq 2$ this construction can be modified to give a kernel with $\mathfrak{R}_m$ as domain of regularity.

BROWN UNIVERSITY

BOUNDED SELF-ADJOINT OPERATORS AND THE PROBLEM OF MOMENTS*

B. A. LENGYEL

It is known that there exists a close connection between the theory of moments and Jacobi matrices on one side, and the theory of self-adjoint operators in Hilbert space on the other. This connection has been thoroughly investigated by Stone in the tenth chapter of his textbook on Hilbert space.† The solution of both the moment problem and the spectral resolution of self-adjoint operators relies on the possibility of representing a class of analytic functions with positive imaginary parts in the upper half-plane by *Stieltjes* integrals of the form

$$\int_{-\infty}^{+\infty} \frac{d\rho(\lambda)}{\lambda - z}.$$

However, the spectral resolution requires the representation of more general functions than those involved in the solution of the problem of moments. The *bounded* self-adjoint operators do not, and I wish to show that the spectral theorem for *bounded* operators can be deduced immediately from the well known theorems concerning the problem of moments. Let H be a bounded self-adjoint operator, f an element of the Hilbert space, $R_z = (H - zI)^{-1}$ the resolvent of H. It

* Presented to the Society, September 6, 1938.

† M. H. Stone, *Linear Transformations in Hilbert Space and their Applications to Analysis*, American Mathematical Society Colloquium Publications, vol. 15, New York, 1932. The notations of this textbook are used throughout the present paper.

is known that in order to prove the standard form of the spectral theorem it is sufficient to prove that given any f there exists a unique monotone increasing function $\rho(\lambda)$ of bounded variation, such that

$$(R_z f, f) = \int_{-\infty}^{+\infty} \frac{d\rho(\lambda)}{\lambda - z}$$

for all values of z.* The resolvent of a *bounded* self-adjoint operator can be developed in a series:

$$-R_z = \frac{I}{z} + \frac{H}{z^2} + \frac{H^2}{z^3} + \cdots,$$

or

$$-(R_z f, f) = \frac{c_0}{z} + \frac{c_1}{z^2} + \frac{c_2}{z^3} + \cdots,$$

where $c_i = (H^i f, f)$, and the series is convergent for sufficiently large values of z.

Now, if the moment problem has a solution for the sequence $\{c_i\}$; that is, if there exists a $\rho(\lambda)$ such that

$$c_i = \int_{-\infty}^{+\infty} \lambda^i d\rho(\lambda)$$

for all i, then

$$(R_z f, f) = -\sum_{i=0}^{\infty} \int_{-\infty}^{+\infty} \frac{\lambda^i}{z^{i+1}} d\rho(\lambda) = \int_{-\infty}^{+\infty} \frac{d\rho(\lambda)}{\lambda - z},$$

and the spectral problem is solved.

Let us consider the quadratic forms

$$\left| \sum_{i=0}^{n} x_i H^i f \right|^2 = \sum_{i,j=0}^{n} x_i \bar{x}_j (H^i f, H^j f) = \sum_{i,j=0}^{n} c_{i+j} x_i \bar{x}_j, \qquad n = 1, 2, \cdots.$$

These forms are all definite positive for every f and n; therefore their determinants

$$c_n = \begin{vmatrix} c_0 & c_1 & \cdots & c_n \\ c_1 & c_2 & \cdots & c_{n+1} \\ \cdot & \cdot & \cdots & \cdot \\ c_n & c_{n+1} & \cdots & c_{2n} \end{vmatrix}$$

* Cf. B. O. Koopman and J. L. Doob, this Bulletin, vol. 40 (1934), pp. 601–605, or B. A. Lengvel. Acta Scientiarum Mathematicarum. Szeged, vol. 9 (1939).

are all positive. This is the well known condition for the existence of a solution of the problem of moments.

In order to prove the uniqueness of the solution we make use of a theorem of Hamburger.* Let M_n be the minimum of the definite quadratic form

$$\sum_{i,j=0}^{n} c_{i+j}x_i\bar{x}_j$$

for $x_0=1$, and let N_n be the minimum of

$$\sum_{i,j=0}^{n} c_{i+j+2}x_i\bar{x}_j$$

for $x_0=1$. Evidently both M_n and N_n are nonnegative; furthermore $M_n \geqq M_{n+1}$, $N_n \geqq N_{n+1}$; therefore, $M=\lim M_n$ and $N=\lim N_n$ exist.

The necessary and sufficient condition for the uniqueness of the solution of the moment problem is that at least one of the numbers M and N is 0.

Thus, to prove the uniqueness of $\rho(\lambda)$ we have to show that for every f, at least one of the sequences

$$M_n = \min \left| f - \sum_{i=1}^{n} x_i H^i f \right|^2, \qquad N_n = \min \left| Hf - \sum_{i=1}^{n} x_i H^{i+1} f \right|^2$$

tends to 0. This means that f can be approximated by linear combinations of $Hf, H^2f, \cdots$; or Hf can be approximated by linear combinations of $H^2f, H^3f, \cdots$; or both approximations are possible. In other words, if $\mathfrak{M}_f$ is the closed linear manifold determined by $Hf, H^2f, \cdots$, then at least one of the following inclusion relations holds:

(1) $$f \in \mathfrak{M}_f,$$

(2) $$Hf \in \mathfrak{M}_{Hf}.$$

This lemma is interesting in itself; its proof can be carried out as follows.

Assume that $\mathfrak{M}_n$ is the closed linear manifold determined by $[Hf, H^2f, \cdots, H^nf]$. Evidently $\mathfrak{M}_n \subset \mathfrak{M}_{n+1}$ and $H\mathfrak{M}_n \subset \mathfrak{M}_{n+1}$. Therefore, $\mathfrak{M}_f=\lim_{n\to\infty}\mathfrak{M}_n$ exists and $H\mathfrak{M}_f \subset \mathfrak{M}_f$. The element f can be re-

* H. Hamburger, Mathematische Annalen, vol. 81 (1920), p. 234. Hamburger uses the same quadratic forms with real variables. This evidently does not affect the minima.

solved into the sum of f_1 and f_2, with f_1 in $\mathfrak{M}_f$ and f_2 in $\mathfrak{H}-\mathfrak{M}_f$, the orthogonal complement of $\mathfrak{M}_f$.

Either $f_2=0$, whence $f \in \mathfrak{M}_f$, or $f_2 \neq 0$. In the latter case we apply H^2 to $f=f_1+f_2$ and multiply by f_2. This gives

$$(H^2f, f_2) = (H^2f_1, f_2) + (H^2f_2, f_2).$$

Since H^2f and H^2f_1 are in $\mathfrak{M}_f$, their scalar products with f_2 give 0; thus

$$(H^2f_2, f_2) = |Hf_2|^2 = 0. \tag{3}$$

The element f_1 is in the closed linear manifold $\mathfrak{M}_f$, so that for suitable $x_1^{(n)}, x_2^{(n)}, \cdots, x_n^{(n)}$ we have

$$\left| f - f_2 - \sum_{i=1}^{n} x_i^{(n)} H^i f \right| = \left| f_1 - \sum_{i=1}^{n} x_i^{(n)} H^i f \right| = \epsilon_n,$$

where $\epsilon_n \to 0$. Applying the bounded operator H to

$$f - f_2 - \sum_{i=1}^{n} x_i^{(n)} H^i f$$

and using (3), we obtain

$$\left| Hf - \sum_{i=1}^{n} x_i^{(n)} H^{i+1} f \right| \to 0, \quad \text{if} \quad n \to \infty;$$

that is, (2) holds. Thereby the proof of our statement is complete.

WORCESTER, MASS.

ON THE DETERMINANT OF AN AUTOMORPH OF A NONSINGULAR SKEW-SYMMETRIC MATRIX

JOHN WILLIAMSON

Let G be the skew-symmetric matrix of order $2n$,

$$G = \begin{pmatrix} 0 & E_n \\ -E_n & 0 \end{pmatrix},$$

where E_n is the unit matrix of order n. If F is a matrix which satisfies

$$FGF' = G, \tag{1}$$

then $|F|^2 = 1$, so that $|F| = \pm 1$. That $|F| = +1$ is well known and is in fact a consequence of a theorem of Frobenius.* A simple proof communicated to me by Professor Wintner depends on the polar factorization of F, which reduces the problem at once to the case in which F is orthogonal. This proof is, of course, not valid in any field. It is our intention here to give a simple direct proof, applicable in any field, of the fact that $|F| = +1$.

On writing F as a matrix of matrices of order n, $F = (F_{ij})$, $(i, j = 1, 2)$, we have, as a consequence of (1),

$$\begin{aligned} F_{11}F'_{12} - F_{12}F'_{11} &= F_{21}F'_{22} - F_{22}F'_{21} = 0, \\ F_{11}F'_{22} - F_{12}F'_{21} &= F_{22}F'_{11} - F_{21}F'_{12} = E_n. \end{aligned} \tag{2}$$

Let $|F_{11}| \neq 0$. Then

$$F = \begin{pmatrix} F_{11} & F_{12}F'_{11} \\ F_{21} & F_{22}F'_{11} \end{pmatrix} \begin{pmatrix} E_n & 0 \\ 0 & (F'_{11})^{-1} \end{pmatrix}.$$

On, applying (2), we have

$$\begin{aligned} |F'_{11}||F| &= \begin{vmatrix} F_{11} & F_{12}F'_{11} \\ F_{21} & F_{22}F'_{11} \end{vmatrix} = \begin{vmatrix} F_{11} & F_{12}F'_{11} - F_{11}F'_{12} \\ F_{12} & F_{22}F'_{11} - F_{21}F'_{12} \end{vmatrix} \\ &= \begin{vmatrix} F_{11} & 0 \\ F_{21} & E_n \end{vmatrix} = |F_{11}|. \end{aligned}$$

---

* G. Frobenius, *Ueber die schiefe Invariante einer bilinearen oder quadratischen Form*, Journal für die reine und angewandte Mathematik, vol. 86 (1879), pp. 44–71; in particular, p. 48. See A. Wintner, *On linear conservative dynamical systems*, Annali di Matematica Pura ed Applicata, (4), vol. 13 (1934–1935), pp. 105–112.

Therefore $|F| = +1$, and we have proved the following lemma:

LEMMA 1. *If F_{11} is nonsingular, $|F| = +1$.*

It also follows from (2) that if $F_{12} = F_{21} = 0$, $F_{11}F'_{22} = E_n$, so that F_{11} is nonsingular and accordingly $|F| = +1$.

Let P_{ij} be the permutation matrix of order n, which by post-multiplication interchanges the ith and the jth columns of a matrix, and let P be the *diagonal block* matrix

$$P = [P_{ij}, P_{ij}] = \begin{pmatrix} P_{ij} & 0 \\ 0 & P_{ij} \end{pmatrix}.$$

Since P_{ij} is symmetric and involutory, $PGP' = G$, the matrix FP satisfies (1), and $|FP| = |F|$. Consequently we have the following lemma:

LEMMA 2. *Any matrix F_1 obtained from F by a permutation of its first n columns and the same permutation of its last n columns also satisfies* (1) *and* $|F_1| = |F|$.

The matrix $W = (W_{ij})$, $(i, j = 1, 2)$, where $W_{11} = W_{22} = [0, E_{n-1}]$ and $W_{12} = -W_{21} = [E_1, 0]$, satisfies (1) and has determinant unity. The matrix FW is obtained from F by replacing the first column by minus the $(n+1)$st column and the $(n+1)$st by the first. If, for convenience, we now write $F_{11} = A$ and $F_{12} = B$ and denote the columns of A and B by a_i and b_i, respectively, we have, as a consequence of Lemma 2, the following lemma:

LEMMA 3. *The matrix $A = F_{11}$ in F may be replaced by $C = (c_1, c_2, \cdots, c_n)$, where $c_i = a_i$ or $-b_i$.*

Therefore by Lemma 1, since $|W| = +1$, we have our fourth lemma:

LEMMA 4. *If there exists a matrix C such that $|C| \neq 0$, then $|F| = +1$.*

Let every determinant of order n formed from (A, B), in which less than r pairs of columns a_i, b_i occur with the same suffix i, be zero, but let at least one determinant with exactly r pairs of columns a_i, b_i be different from zero. As a consequence of Lemma 2 there is no loss in generality in assuming that

$$|a_1 b_1 X| \neq 0, \tag{3}$$

where the matrix X contains exactly $r-1$ pairs of columns a_i, b_i with the same subscript i and does not contain either of the columns a_2 or b_2. Let Q be the diagonal block matrix

$$Q = \left[\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, E_{n-1}, \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}, E_{n-2}\right].$$

Then $|Q| = +1$, $QGQ' = G$, and FQ satisfies (1). The matrix of the first n rows of FQ is (H, K), where the n columns h_j of H are given by

$$h_j = a_j, \quad j \neq 2, \qquad h_2 = a_1 + a_2, \tag{4}$$

and the n columns k_i of K by

$$k_i = b_i, \quad i \neq 1, \qquad k_1 = b_1 - b_2. \tag{5}$$

Since the matrix X in (3) does not contain any of the columns a_1, b_1, a_2, b_2, it follows from (4) and (5) that the matrix $T = (h_2 k_1 X)$ is a submatrix of (H, K), which contains exactly $r-1$ pairs of columns h_i, k_i with the same suffix i, and that

$$\begin{aligned} |T| &= |a_1 + a_2, b_1 - b_2, X| \\ &= |a_1 b_1 X| - |a_1 b_2 X| + |a_2 b_1 X| - |a_2 b_2 X|. \end{aligned} \tag{6}$$

But, by hypothesis,

$$|a_1 b_2 X| = |a_2 b_1 X| = 0. \tag{7}$$

Since $|a_1 a_2 X|$ is also zero and $|a_1 b_1 X|$ is not zero by (3), a_2 and b_2 are both linear combinations of the $n-1$ columns of the matrix $(a_1 X)$. Hence $|a_2 b_2 X| = 0$ and, as a consequence of (6) and (7), $|T| = |a_1 b_1 X| \neq 0$.

Therefore in (H, K) there is one nonzero subdeterminant of order n which contains exactly $r-1$ pairs of columns h_i, k_i with the same suffix i.

Now $|FQ| = |F|$, and FQ satisfies (1). Further, the matrix C in Lemma 4 contains exactly $r = 0$ pairs of columns a_i, b_i. By a simple induction proof we therefore have the following lemma:

LEMMA 5. *If in (A, B) there is one nonzero subdeterminant of order n which contains $r \leqq n$ pairs of columns a_i, b_i with the same suffix i, then $|F| = +1$.*

Since any matrix F which satisfies (1) is nonsingular, the rank of (A, B) is n and Lemma 5 implies the following statement:

THEOREM. *If $FGF' = G$, $|F| = +1$.*

This proof is valid in any field.

THE JOHN HOPKINS UNIVERSITY

ON GREEN'S FUNCTIONS IN THE THEORY OF HEAT CONDUCTION IN SPHERICAL COORDINATES†

ARNOLD N. LOWAN

In a previous paper,‡ the writer derived the expressions for the Green's functions in the theory of heat conduction for an infinite cylinder and for an infinite solid, bounded internally by a cylinder.

The object of the present paper is to derive the appropriate Green's functions for a sphere and for an infinite solid bounded internally by a sphere. In both cases, we shall take the boundary condition in the form

$$\frac{\partial u}{\partial r} + hu = 0, \qquad r = a.$$

The case of a sphere. In this case we start with the expression

$$u(r, \theta, \phi, t; r_0, \theta_0, \phi_0) = \frac{1}{2(\pi kt)^{3/2}} e^{-R^2/4kt}, \tag{1}$$

where

$$R^2 = r^2 + r_0^2 - 2r_0 \cos \gamma, \tag{2}$$

γ being the angle between the radii from the origin to the points (r, θ, ϕ) and (r_0, θ_0, ϕ_0). The expression (1) is the point source solution of the differential equation of heat conduction in spherical coordinates.

The expression (1) may be written in the form§

$$u(r, \theta, \phi, t; r_0, \theta_0; \phi_0) = \frac{1}{4\pi(rr_0)^{1/2}} \sum_{n=0}^{\infty} (2n + 1) P_n(\cos \gamma) \int_0^{\infty} \alpha e^{-k\alpha^2 t} J_{n+1/2}(\alpha r_0) J_{n+1/2}(\alpha r) d\alpha. \tag{3}$$

The corresponding Laplace transform

$$L\{u(t)\} = \int_0^{\infty} e^{-pt} u(t) dt = u^*(p)$$

† Presented to the Society, October 29, 1938.

‡ This Bulletin, vol. 44 (1938), pp. 125–133. This paper will be referred to as A.N.L.

§ See Carslaw, *Mathematical Theory of Heat Conduction*, article 93.

is therefore

$$u^*(r, \theta, \phi, p; r_0, \theta_0, \phi_0) = \frac{1}{4\pi(rr_0)^{1/2}} \sum_{n=0}^{\infty} (2n+1) P_n(\cos\gamma) \cdot \int_0^{\infty} \frac{\alpha d\alpha}{\alpha^2 - q^2} J_{n+1/2}(\alpha r) J_{n+1/2}(\alpha r_0), \tag{4}$$

where we have put $p = -kq^2$.

With the aid of the identities (5) and (5′) of A.N.L., (4) becomes

$$u^* = \frac{i}{8k(rr_0)^{1/2}} \sum_{n=0}^{\infty} (2n+1) P_n(\cos\gamma) J_{n+1/2}(rq) H^1_{n+1/2}(r_0 q), \qquad r < r_0, \tag{5}$$

$$u^* = \frac{i}{8k(rr_0)^{1/2}} \sum_{n=0}^{\infty} (2n+1) P_n(\cos\gamma) J_{n+1/2}(r_0 q) H^1_{n+1/2}(rq), \qquad r > r_0. \tag{6}$$

In order to obtain the Green's function, we must add to the point source solution u a function v, satisfying the differential equation of heat conduction, vanishing at $t=0$, and such that $u+v$ satisfies the boundary condition $\partial u/\partial r + hu = 0$, for $r=a$.

Since $L\{\partial u(t)/\partial t\} = L\{u(t)\} - u(0) = u^*(p) - u(0)$, and since the two operations of differentiation with respect to x, and of acting with the Laplace operator L, may be commuted, the Laplace transform of v must satisfy the differential equation

$$\Delta v^* + q^2 v^* = 0. \tag{7}$$

The transition from u^*+v^* to the desired Green's function $G = u+v$. will be apparent from the subsequent developments.

The most general solution of (7) which is symmetric about the axis $\gamma = 0$ may be written in the form

$$v^* = \frac{i}{8k(rr_0)^{1/2}} \sum_{n=0}^{\infty} (2n+1) A_n P_n(\cos\gamma) J_{n+1/2}(rq). \tag{8}$$

From (8) we get

$$\left(\frac{\partial v^*}{\partial r} + hv^*\right)_{r=a} = \frac{i}{8k(ar_0)^{1/2}} \sum_{n=0}^{\infty} (2n+1) A_n P_n(\cos\gamma) \cdot \left\{ q \frac{d}{dz} J_{n+1/2}(z) + \left[h - \frac{1}{2a}\right] J_{n+1/2}(z) \right\}_{z=aq}. \tag{9}$$

Since

$$\left(\frac{\partial}{\partial r} + h\right)(u^* + v^*) = 0, \qquad r = a, \tag{10}$$

it follows that

$$A_n = -J_{n+1/2}(r_0 q)\frac{\left\{q\frac{d}{dz}H^1_{n+1/2}(z) + (h - 1/(2a))H^1_{n+1/2}(z)\right\}_{z=aq}}{qJ'_{n+1/2}(aq) + (h - 1/(2a))J_{n+1/2}(aq)}; \tag{11}$$

therefore†

$$u^* + v^* = \frac{i}{8k(rr_0)^{1/2}}\sum_{n=0}^{\infty}(2n+1)P_n(\cos\gamma)w_n^*, \tag{12}$$

where

$$w_n^* = \frac{J_{n+1/2}(rq)}{U_{n+1/2}(aq)}\left\{H^1_{n+1/2}(rq)U_{n+1/2}(aq) - J_{n+1/2}(rq)\cdot\left[\frac{z}{a}\frac{d}{dz}H^1_{n+1/2}(z) + \left(h - \frac{1}{2a}\right)H^1_{n+1/2}(z)\right]_{z=aq}\right\} \tag{13}$$

and

$$U_{n+1/2}(aq) = qJ'_{n+1/2}(aq) + \left(h - \frac{1}{2a}\right)J_{n+1/2}(aq). \tag{14}$$

Comparison between (14) and equation (14) of A.N.L. shows clearly that there is a formal analogy between the present and the former expression for w_n^*. Specifically, our present w_n^* may be obtained from the corresponding expression in A.N.L. by replacing n by $n+1/2$ and h by $h-1/(2a)$ and multiplying by the factor $1/2$. The inversion of (12) therefore ultimately yields‡

$$G(r,\theta,\phi,t;r_0,\theta_0,\phi_0) = \frac{1}{2\pi a^2(r\,r_0)^{1/2}}\sum_{n=0}^{\infty}(2n+1)P_n(\cos\gamma)\cdot\sum_{q_i}q_i^2 e^{-kq_i^2 t}J_{n+1/2}(q_i r)\cdot\frac{J_{n+1/2}(q_i r_0)}{\left[(h-1/(2a))^2 + q_i^2 - (n+1/2)^2/a^2\right]\left[J_{n+1/2}(q_i a)\right]^2}, \tag{15}$$

† Formulas (13), (15), (18), (19), (20), and (22) are given for $r<r_0$. In the case $r>r_0$, the corresponding formulas are obtained by interchanging r and r_0.

‡ As mentioned in A.N.L., the transition from $pw_n^* = Y(p)/Z(p)$ to w_n is equivalent to the inversion of the Laplace transform defining w_n^*, and we have

$$w_n = \frac{Y(0)}{Z(0)} + \sum\frac{Y(p_i)}{p_i Z'(p_i)}\cdot e^{p_i t},$$

where the summation extends over the roots of $Z(p)=0$.

where the second summation extends over the roots

$$U_{n+1/2}(aq) = 0. \tag{16}$$

From this formula we may obtain the Green's function for the case where the boundary is impervious to heat by putting $h=0$. Also the case where the boundary is kept at 0° may be obtained by putting $h=\infty$. In this case it is clear that the transcendental equation (16) reduces to

$$J_{n+1/2}(aq) = 0. \tag{17}$$

Also it is easily seen that the denominator of (15) reduces to

$$q_i^2 [J'_{n+1/2}(q_i a)]^2.$$

Thus the Green's function for the case where the boundary is kept at 0° is

$$G(r, \theta, \phi, t; r_0, \theta_0, \phi_0) = \frac{1}{2\pi a^2 (r r_0)^{1/2}} \sum_{n=0}^{\infty} (2n+1) P_n(\cos\gamma) \sum_{q_i} e^{-kq_i^2 t} \cdot \frac{J_{n+1/2}(q_i r) J_{n+1/2}(q_i r_0)}{\{J'_{n+1/2}(aq_i)\}^2}, \tag{18}$$

where the second summation extends over the roots of (17).

Case of the infinite solid bounded internally by a sphere. The former analogy with the treatment in A.N.L., noticed in the previous case, applies also in the case under consideration. Thus since v^* must be finite for $r=\infty$, it follows that in (8) we must replace $J_{n+1/2}(rq)$ by $H^1_{n+1/2}(rq)$. Proceeding as in the previous case, we ultimately obtain

$$u^* + v^* = \frac{i}{8k(rr_0)^{1/2}} \sum_{n=0}^{\infty} (2n+1) P_n(\cos\gamma) W_n^*, \tag{19}$$

where the expression for W_n^* may be obtained from equation (30) of A.N.L. by replacing h by $h-1/(2a)$ and n by $n+1/2$ and multiplying by the factor $1/2$. Our final solution is therefore

$$G(r, \theta, \phi, t; r_0, \theta_0, \phi_0) = \frac{1}{8\pi(rr_0)^{1/2}} \sum_{n=0}^{\infty} (2n+1) P_n(\cos\gamma) \cdot \int_{-\infty}^{+\infty} \alpha e^{-k\alpha^2 t} \frac{H^1_{n+1/2}(\alpha r_0)}{U_{n+1/2}(\alpha a)} \cdot \{J_{n+1/2}(\alpha r) U_{n+1/2}(\alpha a) - U_{n+1/2}(\alpha r) J_{n+1/2}(\alpha a)\} d_\alpha \tag{20}$$

where

$$U_{n+1/2}(\alpha a) = \left\{\alpha \frac{d}{dz} H^1_{n+1/2}(z) + \left(h - \frac{1}{2a}\right) H^1_{n+1/2}(z)\right\}_{z=\alpha a}. \tag{21}$$

For $h = \infty$ this reduces to

$$\begin{aligned} G(r, \theta, \phi, t; r_0, \theta_0, \phi_0) = {} & \frac{1}{8\pi(rr_0)^{1/2}} \sum_{n=0}^{\infty} (2n+1) P_n(\cos\gamma) \\ & \cdot \int_{-\infty}^{+\infty} \alpha e^{-k\alpha^2 t} \frac{H^1_{n+1/2}(\alpha r_0)}{H^1_{n+1/2}(\alpha a)} \\ & \cdot \{J_n(\alpha r) H^1_{n+1/2}(\alpha a) - J_{n+1/2}(\alpha a) H^1_{n+1/2}(\alpha r)\} d\alpha. \end{aligned} \tag{22}$$

This is the solution of our problem when the spherical surface $r = a$ is kept at 0°.

The Green's functions above evaluated may be called *point source Green's functions*. They are solutions of the differential equation of heat conduction, depending on the spherical coordinates r, θ, and ϕ and satisfying the condition

$$\lim_{\epsilon \to 0} \iiint_{\omega} G(r, \theta, \phi, 0; r', \theta', \phi') d\tau = 1, \tag{23}$$

where ω is a little sphere of radius ϵ surrounding the point source (r_0, θ_0, ϕ_0).

In addition to these Green's functions we may consider the Green's functions depending on r only and satisfying the condition

$$\lim_{\epsilon \to 0} 4\pi \int_{r_0}^{r_0+\epsilon} G(r, \rho, 0) \rho^2 d\rho = 1. \tag{24}$$

For the case of the sphere radiating into a medium at 0°, the Green's function, while not given explicitly by Carslaw, may be derived from his article 65, in the form

$$\begin{aligned} G(r, t; r_0) = {} & \frac{1}{2\pi a r r_0} \sum_{n=1}^{\infty} \frac{a^2\alpha_n^2 + (ah-1)^2}{a^2\alpha_n^2 + ah(ah-1)} \\ & \cdot \sin \alpha_n r \cdot \sin \alpha_n r_0 e^{-k\alpha_n^2 t}, \end{aligned} \tag{25}$$

where α_n is a root of $a\alpha \cos a\alpha + (ah-1) \sin a\alpha = 0$.

The Green's function for the case of the infinite solid bounded internally by a sphere may be obtained by considering a continuous distribution of point sources over the sphere $r = r_0$ and integrating for the variables θ' and ϕ'. This leads to

$$G(r, t; r_0) = \frac{1}{2\pi a^2} \sum_{q_i} P_0(\cos \gamma) e^{-kq_i^2 t} \cdot \frac{J_{1/2}(q_i r) J_{1/2}(q_i r_0)}{[(h - 1/(2a))^2 + q_i^2 - 1/(4a^2)][J_{1/2}(q_i a)]^2}, \tag{26}$$

where the summation extends over the roots of (17).

The desired results may also be obtained in the following manner. It can be easily shown that if $u(r, \rho, t)$, is the Green's function appropriate to a "plane source," and therefore satisfying the condition

$$\lim_{\epsilon \to 0} \int_{r_0}^{r_0+\epsilon} u(r, \rho, 0) d\rho = 1, \tag{27}$$

then

$$v = \frac{1}{4\pi r r_0} u$$

is the desired Green's function appropriate to a spherical source. By substituting for u the expression which may be derived from Carslaw's article 82, the desired Green's function is obtained in the form

$$G(r, t; r_0) = \frac{1}{8\pi r r_0 (\pi k t)^{1/2}} \left\{ \exp\left[-\frac{(r - a - r_0)^2}{4kt} \right] + \exp\left[-\frac{(r - a + r_0)^2}{4kt} \right] - 2h \int_0^\infty e^{-h\xi} \exp\left[-\frac{(r - a + r_0 + \xi)^2}{4kt} \right] d\xi \right\} \tag{28}$$

which must, of course, agree with (26).

It should be remarked that the Green's functions so derived are of the general form

$$G = \sum u_n(P) \cdot u_n(P_0) e^{-k\lambda_n^2 t}, \tag{29}$$

where the u_n's are the normalized characteristic solutions of the homogeneous differential equation of

$$\nabla^2 u + \lambda^2 u = 0 \tag{30}$$

which satisfies the prescribed boundary conditions.

BROOKLYN COLLEGE AND
YESHIVA COLLEGE

ON WAVE MOTION IN AN INFINITE SOLID BOUNDED INTERNALLY BY A CYLINDER OR A SPHERE

ARNOLD N. LOWAN

PART I

In two previous papers,† the author investigated the problem of wave motion for infinite domains of one, two, and three dimensions and for certain sub-infinite domains; that is, domains bounded in certain directions but extending to infinity in other directions. The present paper is a sequel to the aforementioned papers and deals with the problem of wave motion in an infinite solid, bounded internally by a cylinder or a sphere.

In the subsequent developments we shall use the following abbreviations:

$$\sigma(\alpha) = (a^2\alpha^2 - k^2)^{1/2}, \qquad s(p, \alpha) = \alpha^2 + (p^2 - k^2)/a^2,$$

where α is a real variable ranging from $-\infty$ to ∞ and p is a complex variable whose real part is positive. We shall also introduce the operators ∇_c, ∇_s, $\sum\int\int\int$, and $\sum\int\int\int\int$ defined as follows:

$$\nabla_c = \frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} - \frac{1}{a^2}(p^2 - k^2) + \frac{1}{r^2}\frac{\partial^2}{\partial\theta^2},$$

$$\nabla_s = \frac{\partial^2}{\partial r^2} + \frac{2}{r}\frac{\partial}{\partial r} + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\,\frac{\partial}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2}{\partial\phi^2} - \frac{1}{a^2}(p^2 - k^2),$$

$$\sum\int\int\int\{F_n(r', \theta', \alpha)\} = \sum_{n=0}^{\infty}(2n+1)\cos n(\theta - \theta')\int_R^{\infty} r'dr' \cdot \int_0^{2\pi} d\theta' \int_{-\infty}^{\infty} \alpha F_n(r', \theta', \alpha)d\alpha,$$

$$\sum\int\int\int\int\{F_n(r', \theta', \phi', \alpha) = \sum_{n=0}^{\infty}(2n+1)P_n(\cos\gamma) \cdot \int_R^{\infty} r'^{3/2}dr' \int_0^{\pi}\sin\theta' d\theta' \int_0^{2\pi} d\phi' \int_{-\infty}^{\infty} \alpha F_n(r', \theta', \phi', \alpha)d\alpha,$$

† *On wave motion for infinite domains*, Philosophical Magazine, (7), vol. 26 (1938), pp. 340–360; *On wave motion for sub-infinite domains*, Philosophical Magazine, (7), vol. 27 (1939), pp. 182–194. These papers will be referred to as L-1 and L-2, respectively.

where P_n is the Legendre polynomial of degree n and γ is the angle between the vectors from the origin to the points (r, θ, ϕ) and (r', θ', ϕ').

Consider first the case of an infinite solid bounded internally by a cylinder. In this case the displacement satisfies the system A consisting of equations (1), (2), (3), and (4):

$$\nabla^2 U(P; t) = 2b \frac{\partial}{\partial t} U(P; t) + \frac{1}{a^2} \frac{\partial^2}{\partial t^2} U(P; t) + \Phi(P; t), \tag{1}$$

$$\lim_{t \to 0} U(P; t) = f(P), \tag{2}$$

$$\lim_{t \to 0} \frac{\partial}{\partial t} U(P; t) = g(P), \tag{3}$$

$$U(P; t) = \phi(\theta; t), \qquad r = R. \tag{4}$$

As in L-1 and L-2, we put

$$U(P; t) = u(P; t) + v(P; t), \tag{5}$$

where $u(P; t)$ satisfies the system B, consisting of (1), (2), (3), and the boundary condition

$$u(P; t) = 0, \qquad r = R, \tag{6}$$

and where $v(P; t)$ satisfies the system C, obtained from A, by replacing $U(P; t)$ by $v(P; t)$ and putting $f(P) = g(P) = \Phi(P; t) = 0$. We proceed to the solution of the systems B and C.

As in L-1 and L-2, the method of solving the systems B and C consists in making the substitution

$$u(r, \theta; t) = e^{-kt} u_1(r, \theta; t), \tag{7}$$

$$\Phi(r, \theta; t) = e^{-kt} \Phi_1(r, \theta; t), \tag{8}$$

$$v(r, \theta; t) = e^{-kt} v_1(r, \theta; t), \tag{9}$$

where $k = ba^2$.

Let B_1 and C_1 designate the systems obtained from the systems B and C, by the substitutions (7), (8), and (9) for the functions u_1, v_1, and Φ_1. The solutions of the last two systems are obtained by operating on systems B and C with the Laplace operator and obtaining the systems B_1^* and C_1^*, for the Laplace transforms $u_1^*(r, \theta; p)$ and $v_1^*(r, \theta; p)$. When the solutions of B_1^* and C_1^* have been obtained, the corresponding solutions of B and C are obtained by acting on the corresponding solutions with the inverse Laplace operator. The system B_1^* is ultimately obtained in the following form:

$$\nabla_c u_1^*(r, \theta; p) = -\frac{1}{a^2}\{pf(r, \theta) + h(r, \theta)\} + \Phi_1^*(r, \theta; p)$$

$$= -F(r, \theta; p) \text{ (say)}, \tag{10}$$

$$h(r, \theta) = g(r, \theta) + kf(r, \theta),$$

$$u_1^*(R, \theta; p) = 0, \tag{11}$$

and the system C_1^* in the form

$$\nabla_c v_1^*(r, \theta; p) = 0, \tag{12}$$

$$v_1^*(R, \theta; p) = \phi_1^*(\theta; p). \tag{13}$$

In order to obtain the solution of B_1^*, we make use of the identity†

$$f(r, \theta) = \frac{1}{4\pi}\sum\int\int\int \{f(r', \theta')G_n(r, r'; \alpha)\} \tag{14}$$

where

$$G_n(r, r'; \alpha) = \frac{H_n^1(\alpha r)}{H_n^1(\alpha R)}\{J_n(\alpha r') \cdot H_n^1(\alpha R) - J_n(\alpha R)H_n^1(\alpha r')\}$$

for $r' < r$; the corresponding expression for $r' > r$, is obtained by interchanging r and r'.

In view of (14), it can be verified that the expression

$$u_1^*(r, \theta; p) = \frac{1}{4\pi}\sum\int\int\int \left\{F(r', \theta'; p) \cdot \frac{G_n(r, r'; \alpha)}{s(p, \alpha)}\right\} \tag{15}$$

is the solution of the system B_1^*.

Bearing in mind the significance of $F(r, \theta; p)$ from (10), (15) becomes

$$u_1^*(r, \theta; p) = u_{1,1}^*(r, \theta; p) + u_{1,2}^*(r, \theta; p) + u_{1,3}^*(r, \theta; p), \tag{16}$$

where

$$u_{1,1}^*(r, \theta; p) = \frac{1}{4\pi a^2}\sum\int\int\int \left\{f(r', \theta') \frac{pG_n(r, r'; \alpha)}{s(p, \alpha)}\right\}, \tag{17}$$

$$u_{1,2}^*(r, \theta; p) = \frac{1}{4\pi a^2}\sum\int\int\int \left\{h(r', \theta') \frac{G_n(r, r'; \alpha)}{s(p, \alpha)}\right\}, \tag{18}$$

$$u_{1,3}^*(r, \theta; p) = -\frac{1}{4\pi}\sum\int\int\int \left\{G_n(r, r'; \alpha) \frac{\Phi_1^*(r', \theta'; p)}{s(p, \alpha)}\right\}. \tag{19}$$

† See Appendix at end of this paper, §1.

The problem now reduces to obtaining the inverse Laplace transforms of (17), (18), and (19). Making use of the identities

$$\int_0^\infty e^{-pt}\cos\lambda t\,dt = \frac{p}{p^2+\lambda^2}, \quad \int_0^\infty e^{-pt}\sin\lambda t\,dt = \frac{\lambda}{p^2+\lambda^2},$$

and of Borel's theorem, we finally get

$$(20)\quad u_{1,1}(r,\theta;t) = \frac{1}{4\pi}\sum\iiint\{f(r',\theta')G_n(r,r';\alpha)\cos\sigma(\alpha)t\},$$

$$(21)\quad u_{1,2}(r,\theta;t) = \frac{1}{4\pi}\sum\iiint\left\{h(r',\theta')G_n(r,r';\alpha)\frac{\sin\sigma(\alpha)t}{\sigma(\alpha)}\right\},$$

$$(22)\quad u_{1,3}(r,\theta;t) = -\frac{a^2}{4\pi}\int_0^t\sum\iiint\left\{\Phi_1(r',\theta';t-\tau)G_n(r,r';\alpha)\frac{\sin\sigma(\alpha)\tau}{\sigma(\alpha)}\right\}d\tau.$$

In the case where the solutions f, g, ϕ, and Φ do not depend on θ, it is clear that the final solution is also independent of θ. In this case, it can be shown (see Appendix, §2) that the identity (14) must be replaced by

$$(14')\quad f(r) = \frac{1}{4\pi}\int_R^\infty r'dr'\int_{-\infty}^\infty \alpha f(r')G_0(r,r';\alpha)d\alpha,$$

where the expression for G_0 is obtained from that for G_n, by replacing the index n by zero.

In view of this result, the solutions for $u_{1,1}$, $u_{1,2}$, and $u_{1,3}$, when these solutions do not depend on θ, may be obtained at once from (20), (21), and (22), by dropping the summation sign and the factor $\cos n(\theta-\theta')$ and replacing the subscript n by zero.

We proceed to the solution of the system C_1^*. The expression

$$(23)\quad v_1^*(r,\theta;p) = \frac{1}{\pi}\sum_{n=-\infty}^{\infty}\int_0^{2\pi}\phi_1^*(\theta';p)\cos n(\theta-\theta')w_n^*(r;p)d\theta',$$

where

$$(24)\quad w_n^*(r;p) = \frac{H_n^1(\alpha r)}{H_n^1(\alpha R)}\ (\text{say}),$$

is a solution of (12), satisfying the boundary conditions (13), provided

$$(25)\quad s(p,\alpha) = 0.$$

If the function $w_n(r,t)$ is defined by

$$\int_0^\infty e^{-pt} w_n(r;t)\,dt = \frac{H_n^1(\alpha r)}{H_n^1(\alpha R)} = w_n^*(r;p) \text{ (say)}, \tag{26}$$

then by Borel's theorem, the inverse Laplace transform of (23) is

$$v_1(r,\theta;t) = \frac{1}{\pi}\sum_{-\infty}^{\infty}\int_0^{2\pi}\int_0^t \phi_1(\theta';t-\tau)w_n(r,\tau)\cos n(\theta-\theta')\,d\theta'\,d\tau. \tag{27}$$

Our problem thus reduces to solving the integral equation (26), where α is defined in (25). Since $H_n^1(z)\to 0$ in the upper half of the z-plane and since its zeros are known to lie in the lower half of the plane, it can be easily shown with the aid of the Cauchy integral theorem that

$$\begin{aligned} w_n^*(r;p) &= \frac{H_n^1(r\alpha)}{H_n^1(R\alpha)} = \frac{1}{\pi i}\int_{-\infty}^{\infty} \frac{x}{x^2-\alpha^2}\cdot\frac{H_n^1(rx)}{H_n^1(Rx)}\,dx \\ &= \frac{a^2}{\pi i}\int_{-\infty}^{\infty} \frac{x}{p^2+[\sigma(x)]^2}\cdot\frac{H_n^1(rx)}{H_n^1(Rx)}\,dx. \end{aligned} \tag{28}$$

The inversion of (26) leads at once to

$$w_n(r;t) = \frac{a^2}{\pi i}\int_{-\infty}^{\infty} \frac{x\sin\sigma(x)t}{\sigma(x)}\,\frac{H_n^1(rx)}{H_n^1(Rx)}\,dx. \tag{29}$$

In (27) and (29), we have the complete solution of the system B_1.

It should be remarked that the expression $w_n(r;t)$ given by (29) is real. Indeed, if in the contribution to the integral in (29) for the range from $-\infty$ to 0 we make the substitution $x=-\xi$, replace once more the variable of integration ξ by x, and furthermore, make use of the well known relation

$$H_n^1(-z) = (-1)^n[H_n^1(z) - 2J_n(z)], \tag{30}$$

(29) becomes

$$\begin{aligned} w_n(r;t) = \frac{a^2}{\pi i}\int_0^{\infty} \frac{x\sin\sigma(x)t}{\sigma(x)}\Bigg\{&\frac{H_n^1(rx)}{H_n^1(Rx)} \\ &- \frac{H_n^1(rx)-2J_n(rx)}{H_n^1(Rx)-2J_n(Rx)}\Bigg\}\,dx. \end{aligned} \tag{31}$$

Since $H_n^1(z) = J_n(z) + iY_n(z)$, the above equation ultimately becomes

$$w_n(r;t) = \frac{2a^2}{\pi}\int_0^{\infty} \frac{x\sin\sigma(x)t}{\sigma(x)}\cdot\frac{J_n(Rx)Y_n(rx)-J_n(rx)Y_n(Rx)}{(J_n(Rx))^2+(Y_n(Rx))^2}\,dx. \tag{32}$$

If the function ϕ is independent of θ, it is clear that v_1 is independ-

ent of θ. In this case, instead of (23), we start with

$$v_1^*(r; p) = \phi_1^*(p) w_0^*(r; p), \tag{23'}$$

where $w_0^*(r; p)$ may be obtained from (28) by replacing the subscript n by zero. The corresponding expression for $w_0(r, t)$ may therefore be obtained from (29) by replacing the subscript n by zero. With this definition of $w_0(r; t)$ we have

$$v_1(r; t) = \int_0^\infty \phi_1(t - \tau) w_0(r; \tau) d\tau.$$

PART II. CASE OF AN INFINITE SOLID BOUNDED INTERNALLY BY A SPHERE

In this case, the displacement $U(P; t)$ must satisfy a system similar to A, except that $\nabla^2 u(P; t)$ is now the Laplacian in spherical coordinates and the boundary condition (4) may assume the more general form

$$U(P; t) = \Omega(\theta, \phi; t), \qquad r = R. \tag{4'}$$

The method of solution is entirely similar to that in Part I. The Laplace transforms of the functions $u_1(P; t)$ and $v_1(P; t)$ obtained from $u(P; t), v(P; t)$ by the substitutions (7), (8), and (9), for the case under consideration, must now satisfy the systems D_1^*, consisting of equations (33) and (34), and E_1^*, consisting of (35) and (36):

$$\begin{aligned} \nabla_s u_1^*(r, \theta, \phi; p) &= -\frac{1}{a^2}\{pf(r, \theta, \phi) + h(r, \theta, \phi)\} + \Phi_1^*(r, \theta, \phi; p) \\ &= -F(r, \theta, \phi; p) \text{ (say)}, \end{aligned} \tag{33}$$

$$u_1^*(R, \theta, \phi; p) = 0, \tag{34}$$

$$\nabla_s v_1^*(r, \theta, \phi; p) = 0, \tag{35}$$

$$v_1^*(r, \theta, \phi; p) = \Omega_1^*(\theta, \phi; p). \tag{36}$$

We proceed to the solution of the system D_1^*. In this case, we make use of the following identity;†

$$f(r, \theta, \phi) = \frac{1}{8\pi r^{1/2}} \sum \int\int\int\int \{f(r', \theta', \phi') G_{n+1/2}(r, r', \alpha)\}. \tag{37}$$

As in Part I, it can be verified that the expression

† The derivation of this identity is discussed in the Appendix, §4. The expression for $G_{n+1/2}(r, r', \alpha)$ may be obtained from that of $G_n(r, r', \alpha)$ by replacing the subscript n by $n+1/2$.

$$(38)\quad u_1^*(r,\theta,\phi;p)=\frac{1}{8\pi r^{1/2}}\sum\int\int\int\int\left\{F(r',\theta',\phi')\frac{G_{n+1/2}(r,r',\alpha)}{s(p,\alpha)}\right\}$$

is a solution of the system D_1^*, provided that the expression for $G_{n+1/2}(r,r',\alpha)$ in (38) is obtained from the expression $G_n(r,r',\alpha)$ by replacing the subscript n by $n+1/2$. The expressions for $u_{1,1}$, $u_{1,2}$, and $u_{1,3}$ may therefore be obtained at once from the corresponding expressions in Part I in the forms

$$(39)\quad u_{1,1}(r,\theta,\phi;t)=\frac{1}{8\pi r^{1/2}}\sum\int\int\int\int\{f(r',\theta',\phi')\cdot G_{n+1/2}(r,r';\alpha)\cos\sigma(\alpha)t\},$$

$$(40)\quad u_{1,2}(r,\theta,\phi;t)=\frac{1}{8\pi r^{1/2}}\sum\int\int\int\int\left\{h(r',\theta',\phi')\cdot G_{n+1/2}(r,r';\alpha)\frac{\sin\sigma(\alpha)t}{\sigma(\alpha)}\right\},$$

$$(41)\quad u_{1,3}(r,\theta,\phi;t)=-\frac{a^2}{8\pi r^{1/2}}\int_0^t\sum\int\int\int\int\left\{\Phi_1(r',\theta',\phi';t-\tau)\cdot G_{n+1/2}(r,r';\alpha)\frac{\sin\sigma(\alpha)\tau}{\sigma(\alpha)}\right\}d\tau.$$

We now proceed to the solution of the system E_1^*. The expression

$$(42)\quad v_1^*(r,\theta,\phi;p)=\sum_{n=0}^{\infty}\int_0^{\pi}\int_0^{2\pi}(2n+1)P_n(\cos\gamma)\cdot\frac{H^1_{n+1/2}(\alpha r)}{H^1_{n+1/2}(\alpha R)}\Omega_1^*(\theta',\phi';p)\sin\theta'd\theta'd\phi'$$

is a solution of E_1^*, provided

$$(43)\quad s(p,\alpha)=0$$

(see Appendix, §3).

If, then, the function $w_{n+1/2}(r,t)$ is defined by

$$(44)\quad \int_0^{\infty}e^{-pt}w_{n+1/2}(r;t)dt=\frac{H^1_{n+1/2}(\alpha r)}{H^1_{n+1/2}(\alpha R)}=w^*_{n+1/2}(r;p)\ (\text{say}),$$

then by Borel's theorem, the inversion of (42) yields

$$(45)\quad v_1(r,\theta,\phi;t)=\sum_{n=0}^{\infty}\int_0^{\pi}\int_0^{2\pi}\int_0^{t}(2n+1)P_n(\cos\gamma)\cdot\Omega_1(\theta',\phi';t-\tau)w_{n+1/2}(r',\tau)d\theta'd\phi'd\tau.$$

Our problem thus reduces to solving the integral equation (44).

This equation is identical with (26) except that the subscript n is replaced by $n+1/2$. The method of solution of (45) is entirely similar to that of (26), and we finally get

$$w_{n+1/2}(r;t) = \frac{a^2}{\pi i}\int_{-\infty}^{\infty} \frac{x \sin \sigma(x)t}{\sigma(x)} \cdot \frac{H^1_{n+1/2}(rx)}{H^1_{n+1/2}(Rx)}\,dx, \tag{46}$$

or (by a transformation similar to that for (29))

$$w_{n+1/2}(r;t) = \frac{2a^2}{\pi}\int_0^{\infty} \frac{x \sin \sigma(x)t}{\sigma(x)} \cdot \frac{J_{n+1/2}(Rx)Y_{n+1/2}(rx) - J_{n+1/2}(rx)Y_{n+1/2}(Rx)}{(J_{n+1/2}(Rx))^2 + (Y_{n+1/2}(Rx))^2}\,dx. \tag{47}$$

The complete solution in the case of an infinite solid bounded internally by a sphere is therefore given by

$$U(P;t) = e^{-ba^2t}(u_{1,1} + u_{1,2} + u_{1,3} + v), \tag{48}$$

where the terms in parentheses are given by (39), (40), (41), and (45).

An important special case is that in which the functions f, g, Ω, Φ do not depend on the angles θ and ϕ. While the solution can be obtained from the previous solution by integrating the variables θ' and ϕ', it is easier to proceed as follows. We have to solve the system of equations

$$\frac{\partial^2 u}{\partial r^2} + \frac{2}{r}\frac{\partial u}{\partial r} = 2b\frac{\partial u}{\partial t} + \frac{1}{a^2}\frac{\partial^2 u}{\partial t^2} + \Phi(r;t), \qquad R < r < \infty, \tag{49}$$

$$\lim_{t\to 0} u(r;t) = f(r), \qquad R < r < \infty, \tag{50}$$

$$\lim_{t\to 0} \frac{\partial}{\partial t} u(r;t) = g(r), \qquad R < r < \infty, \tag{51}$$

and

$$U(R;t) = \phi(t).$$

If we make the substitution $u(r;t) = (1/r)v(r;t)$, it is readily seen that the function $v(r;t)$ must satisfy the system

$$\frac{\partial^2 v}{\partial r^2} = 2b\frac{\partial v}{\partial t} + \frac{1}{a^2}\frac{\partial^2 v}{\partial t^2} + r\Phi(r;t), \tag{52}$$

$$\lim_{t\to 0} v(r;t) = rf(r), \tag{53}$$

$$\lim_{t\to 0} \frac{\partial}{\partial t} v(r; t) = rg(r), \tag{54}$$

$$V(R; t) = R\phi(t). \tag{55}$$

The system satisfied by $v(r, t)$ is formally identical with that corresponding to wave motion in a semi-infinite solid extending from $r=R$ to $r=\infty$.

The solution corresponding to a solid extending from 0 to ∞ is given in L-2. It is clear that our present solution may be obtained by replacing x by $r-R$, in L-2.

Appendix

1. **Derivation of the identity** (14). Consider the problem of heat conduction in an infinite solid, bounded internally by a cylinder, the surface of which is kept at 0°. The solution of this problem may be obtained with the aid of the appropriate Green's function for a point source. The expression for the Green's function is

$$G(r, \theta; t; r', \theta') = \frac{1}{4\pi} \sum_{-\infty}^{\infty} \cos n(\theta - \theta') \int_{-\infty}^{\infty} \alpha e^{-k\alpha^2 t} \cdot \frac{H_n^1(\alpha r')}{H_n^1(\alpha R)} \{J_n(\alpha r) H_n^1(\alpha R) - J_n(\alpha R) H_n^1(\alpha r)\} d\alpha \tag{I}$$

for $r<r'$.

In the case $r>r'$, the corresponding expression may be obtained by interchanging r and r'. With the aid of the general formulas of Carslaw's† article 80, the solution of the problem of heat conduction under consideration is seen to be

$$T(r, \theta; t) = \frac{1}{4\pi} \sum \iiint \{f(r', \theta') P_n(\cos \gamma) \cdot e^{-k\alpha^2 t} G_n(r, r', \alpha)\}. \tag{II}$$

Putting $t=0$, we obtain the identity (14).

2. **Derivation of identity** (14′). The expression for the Green's function corresponding to a cylindrical source may be obtained by considering a continuous distribution of line sources around the cylinder $r=r'$, integrating for the variable θ' and dividing by 2π.

The corresponding solution of the problem in heat conduction can then be obtained from (II) by dropping the summation sign and the factor $\cos n(\theta-\theta')$ and replacing the subscript n by zero. If in this final solution we make $t=0$, we obtain the identity (14′).

† Carslaw, *Mathematical Theory of Heat Conduction.*

3. **Derivation of identity** (42). Starting with the well known expansion

$$F(\theta,\phi)=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\int_0^{\pi}\int_0^{2\pi}\frac{(n-|m|)!}{(n+|m|)!}\cdot F(\theta',\phi')P_n^m(\cos\theta)P_n^m(\cos\theta')e^{im(\phi-\phi')}\sin\theta'\,d\theta'\,d\phi', \tag{56}$$

where the P_n^m's are the associated Legendre polynomials, and making use of the identity (see Carslaw's article 93)

$$P_n(\cos\gamma)=P_n(\cos\theta)P_n(\cos\theta')+2\sum_{m=1}^{\infty}\frac{(n-|m|)!}{(n+|m|)!}P_n^m(\cos\theta)\cdot P_n^m(\cos\theta')\cos m(\phi-\phi'), \tag{57}$$

we find that the second member of (42) reduces to $\Omega_i^*(\theta,\phi;p)$ for $r=R$.

4. **Derivation of identity** (37). The Green's function corresponding to a point source in an infinite solid bounded internally by a sphere, the surface of which is kept at 0°, is given by

$$G(r,\theta,\phi;t;r_0,\theta_0,\phi_0)=\frac{1}{8\pi(rr_0)^{1/2}}\sum_{n=0}^{\infty}(2n+1)P_n(\cos\gamma)\cdot\int_{-\infty}^{\infty}\alpha e^{-k\alpha^2 t}\frac{H^1_{n+1/2}(\alpha r_0)}{H^1_{n+1/2}(\alpha R)}\cdot\left\{J_n(\alpha r)H^1_{n+1/2}(\alpha R)-J_{n+1/2}(\alpha R)H^1_{n+1/2}(\alpha r)\right\}d\alpha. \tag{58}$$

With the aid of the general formula of Carslaw's article 80, the solution of the appropriate problem in heat conduction is found to be

$$T(r,\theta,\phi;t)=\frac{1}{8\pi r^{1/2}}\sum\int\int\int\int\left\{e^{-k\alpha^2 t}f(r',\theta',\phi')G_{n+1/2}(r,r')\right\}. \tag{59}$$

Putting $t=0$, we obtain identity (37).

BROOKLYN COLLEGE AND
YESHIVA COLLEGE

FUNDAMENTAL REGIONS FOR THE SIMPLE GROUP OF ORDER 60 IN S_4

R. R. STOLL

Introduction. By a finite collineation group of order g, a point will be transformed into g points at the most. A set of points which are permuted among themselves by the operators of the group are said to form a conjugate set of points. A fundamental region in the plane for the group is defined as a system of points which contains one and only one point from every conjugate set. Therefore, for a collineation group of order g, there are g fundamental regions.

For example, consider the group of order four generated by the transformations $x=y', y=-x'$. These transformations consist of the rotations of a point through one, two, three, and four right angles about the origin in the xy-plane. Any two perpendicular lines through the origin divide the plane into four parts which are the fundamental regions for the group on the plane. Points on the two lines themselves may be assigned to either of the two adjacent regions, provided that no point and its transform be put in the same fundamental region. This example demonstrates that the solution of such a problem is not unique.

The definition of fundamental regions is readily extended to groups in more than two variables, and in the case of three variables the problem has received some attention. J. W. Young* obtained a solution for cyclic groups, W. I. Miller† for the G_{168}, and H. F. Price‡ for the G_{24} and the G_{60} when written with real coefficients in every transformation. This paper is devoted to the determination of fundamental regions for the G_{60} when written with complex coefficients in its transformations. The method used is similar to that used by Miller for the G_{168}.

Determination of fundamental regions. Considered as a collineation group, the generating substitutions for the G_{60} are taken to be§

---

* J. W. Young, *Fundamental regions for cyclic groups of linear fractional transformations on two complex variables*, this Bulletin, vol. 17 (1910–1911), pp. 340–344.

† W. I. Miller, *Fundamental regions for the simple group of order* 168 *in* S_4, American Journal of Mathematics, vol. 56 (1934), pp. 316–318.

‡ H. F. Price, *Fundamental regions for certain finite groups in* S_4, American Journal of Mathematics, vol. 40 (1919), pp. 108–113.

§ H. F. Blichfeldt, *Finite Collineation Groups*, University of Chicago Press, 1917, chap. 5, p. 114.

$$S:\ x_0 = x_0',\qquad x_1 = \omega^4 x_1',\qquad x_2 = \omega x_2',$$

$$T:\ \begin{aligned} x_0 &= 1/5^{1/2}(x_0' + x_1' + x_2'),\quad x_1 = 1/5^{1/2}(2x_0' + bx_1' + ax_2'),\\ x_2 &= 1/5^{1/2}(2x_0' + ax_1' + bx_2'),\end{aligned}$$

where ω is the principal fifth root of unity, $a=\omega+\omega^4$, $b=\omega^2+\omega^3$.

Let the homogeneous complex variables x_0, x_1, x_2 be represented by real points in the S_4 of x, y, u, v where $x_0/x_2=x+iu$, $x_1/x_2=y+iv$. It is understood that when the variables are transformed by an operator of the group, the conjugate imaginary variables $\bar{x}_0$, $\bar{x}_1$, $\bar{x}_2$ are transformed by the conjugate imaginary operator.

Consider the six positive forms $A_i=f_i\bar{f}_i$, $(i=0, 1, \cdots, 5)$, where $\bar{f}_i$ is the conjugate imaginary of f_i and $f_0=5^{1/2}x_0$, $f_i=x_0+\omega^{1-i}x_1+\omega^{i-1}x_2$, $(i=1, 2, \cdots, 5)$. These forms are permuted by S and T:

$$S:\ A_1A_2A_3A_4A_5\cdot,\qquad T:\ A_0A_1\cdot A_2A_5\cdot.$$

When writing the G_{60} as a permutation group of degree six, we may choose as the generating substitutions $S=(A_1A_2A_3A_4A_5)$ and $T=(A_0A_1)(A_2A_5)$.

Forming the differences of the forms $A_0, \cdots, A_5$ gives 15 indefinite forms B_{ij}, $(i=1, 2, 3;\ j=0, 1, \cdots, 4)$, where $B_{10}=A_0-A_1$, $B_{20}=A_1-A_2$, $B_{30}=A_1-A_3$, and the others are selected so that S permutes them in the order $B_{i,j}$, $B_{i,j+1}$. For example,

$$B_{10} = 4x_0\bar{x}_0 - x_0\bar{x}_1 - x_0\bar{x}_2 - x_1\bar{x}_0 - x_1\bar{x}_1 - x_1\bar{x}_2 - x_2\bar{x}_0 - x_2\bar{x}_1 - x_2\bar{x}_2.$$

These 15 forms are used to determine the fundamental regions for the G_{60}. In order to change from the homogeneous complex variables x_0, x_1, x_2 and their conjugate imaginaries to the nonhomogeneous real variables x, y, u, v in one of the forms, it is necessary to divide the form by $x_2\bar{x}_2$. If we perform the division on a form and transform the result by an operator that does not leave x_2 fixed, the denominator will be positive and may be replaced by $x_2\bar{x}_2$ without altering the sign of the form. For certain values of the variables one or more of the denominators will vanish; however, they cannot all vanish at once, since the variables x_0, x_1, x_2 represent homogeneous coordinates of points in a plane.

Dividing each form by $x_2\bar{x}_2$ and replacing the fractions by their equivalents in terms of x, y, u, v gives, if each result is equated to zero, 15 hypersurfaces in S_4. If we exclude points on one or more of the hypersurfaces, a point of S_4 will make each of these forms either positive or negative. There will be at most 6! possible arrangements of sign as these 15 forms are the differences of six positive forms. Later it is shown that all of the 6! arrangements are present. An arrange-

ment of signs is most conveniently given by writing the order of magnitude of the six forms $A_0, \cdots, A_5$; and if $A_0>A_1>A_2>A_3>A_4>A_5$, we shall denote this by $A_0A_1A_2A_3A_4A_5$. This means $B_{10}>0$, $B_{20}>0$, and so on. If we assume, for the present, that every value system (that is, every arrangement of magnitude) of $A_0, \cdots, A_5$ is possible, it follows that S_4 is divided into 6! regions by the 15 forms, and no operator of the group except the identity transforms a point of one region into a point of the same region, since no operator on $A_0, \cdots, A_5$, except the identity, leaves a given arrangement of letters unchanged. Thus each region has 60 conjugates under the group; and since $6!=60\times12$, there are 12 such sets of conjugates. To obtain a fundamental region we select one region from each of the 12 sets.

Let us formulate a scheme to determine to which one of the 12 sets, which we shall denote by S_{ij}, $(i=1, 2; j=1, 2, \cdots, 6)$, a given value system of $A_0, \cdots, A_5$ belongs. The following rules serve to classify the different value systems:

(a) If the first letter is not A_0 (say A_k), transform the value system by a suitable operator of G_{60} so that A_k is replaced by A_0. The operator $S^{-i}TS^i$, which represents

$$(A_0A_1)(A_2A_5), \quad (A_0A_2)(A_1A_3), \quad (A_0A_3)(A_2A_4,)$$
$$(A_0A_4)(A_3A_5), \quad (A_0A_5)(A_1A_4)$$

for $i=0, 1, 2, 3, 4$, serves this purpose.

(b) If the second letter is not A_1, transform by an operator of the G_{60} that omits A_0 and replaces the second letter by A_1. A power of $S=(A_1A_2A_3A_4A_5)$ will accomplish this.

(c) If the third letter is neither A_2 nor A_3, transform by $S^2TS(S^2T)^2=(A_2A_5)(A_3A_4)$; then our value system will have one of the two forms $A_0A_1A_2xxx$, $A_1A_2A_3xxx$, where the x's denote any of the remaining letters. The first subscript is determined as follows: S_{1j} contains $A_0A_1A_2xxx$; S_{2j} contains $A_0A_1A_3xxx$.

(d) The second subscript is determined by the arrangement of the remaining three letters in S_{ij}. The symbol A_p denotes either A_2 or A_3:

Order of A_p, A_4, A_5	$A_pA_4A_5$	$A_pA_5A_4$	$A_4A_pA_5$	$A_5A_pA_4$	$A_4A_5A_p$	$A_5A_4A_p$
Second subscript	1	2	3	4	5	6

Example. $A_1A_3A_5A_0A_2A_4$ is transformed into $A_0A_1A_2A_3A_4A_5$ by the above rules; hence it belongs to S_{11}.

It is interesting to study the intersections of the fundamental regions with certain planes in S_4; in addition this enables us to prove

the existence of the 12 sets of regions S_{ij} by exhibiting points in each set. We shall mention only two planes. (i) Consider the plane $x=0$, $y=0$. The hypersurfaces each intersect this plane in a distinct curve. Included in the intersections are 13 hyperbolas and two straight lines. Upon computing the value systems for the regions intersected by this plane we find that regions from each of the 12 sets S_{ij} are represented. (ii) Consider the plane $u=0$, $v=0$. The forms B_{22} and B_{24} vanish for every point of this plane; it is therefore part of a boundary region. Each region of the plane is on the boundary of four regions in S_4.

The existence of the 12 sets of regions S_{ij} with 60 regions to a set means that it is possible to obtain points whose coordinates give the six forms $A_0, \cdots, A_5$ any preassigned order of magnitude, excluding cases in which two or more of the forms are equal. As stated before, a fundamental region is obtained by selecting one value system for $A_0, \cdots, A_5$ from each of the 12 sets; all points whose coordinates arrange these forms in any one of the 12 value systems thus selected belongs to the same fundamental region. Each fundamental region has 60 conjugates under the group.

As an example of a fundamental region we take the 12 value systems into which all others are transformed by the rules stated above. These are the value systems for which

$$A_0 > A_1 > A_2, \qquad A_1 > A_3, \qquad A_1 > A_4, \qquad A_1 > A_5;$$

and, furthermore, the quantity next in order of magnitude to A_1 is either A_2 or A_3. To complete the determination of the fundamental region given above we assign to it suitable points on the hypersurfaces $B_{ij}=0$. We start with the 12 value systems in which the forms $A_0, \cdots, A_5$ are all distinct and set two or more of the forms equal in all possible ways. After discarding duplicates, we arrange these value systems into sets of conjugates under the group and from each conjugate set select one value system. The value systems thus obtained, together with the 12 in which no two forms are equal, give a complete determination of the fundamental region.

RENSSELAER POLYTECHNIC INSTITUTE

DIOPHANTINE EQUATIONS WHOSE MEMBERS ARE HOMOGENEOUS*

A. A. AUCOIN AND W. V. PARKER

Desboves† has stated that a necessary and sufficient condition for the equation $ax^m+by^m=cz^n$ to have a solution in integers is that c be of the form as^m+bt^m. This would seem to imply‡ that the values of c are restricted whatever be the values of m and n. That this is not the case follows from our first theorem:

THEOREM 1. *If f and g are homogeneous polynomials with integral coefficients, of degrees m and n, respectively, where m and n are relatively prime, then the equation*

$$f(x_1, x_2, \cdots, x_r) = g(y_1, y_2, \cdots, y_s) \tag{1}$$

always has solutions in integers x_i and y_j; and every solution in which the members of (1) *do not vanish is equivalent (in a sense to be defined) to one of the infinitude of solutions given by*

$$x_i = \alpha_i[f(\alpha)]^{n-p}[g(\beta)]^p, \qquad y_j = \beta_j[f(\alpha)]^{m-q}[g(\beta)]^q, \tag{2}$$

where α_i and β_j are arbitrary integers, p and q are integers defined by

$$0 \leqq p \leqq n, \qquad 0 \leqq q \leqq m, \qquad mp - nq = 1, \tag{3}$$

and $f(\alpha)=f(\alpha_1, \alpha_2, \cdots, \alpha_r)$, $g(\beta)=g(\beta_1, \beta_2, \cdots, \beta_s)$.

Since m and n are relatively prime, there exist integers p and q§ such that $0\leqq p\leqq n$, $0\leqq q\leqq m$, $mp=nq+1$, and consequently $n(m-q)=m(n-p)+1$.

If in (1) we let

$$x_i = \alpha_i t^p u^{n-p}, \qquad y_j = \beta_j t^q u^{m-q}, \tag{4}$$

we have‖

---

* Presented to the Society, April 15, 1939.

† A. Desboves, *Mémoire sur la résolution en nombres entières de l'équation $ax^m+by^m=cz^n$*, Nouvelles Annales de Mathématiques, (2), vol. 18 (1879), p. 481. An examination of Desboves's proof shows that he really means that c multiplied by the nth power of an integer u must be of the form as^m+bt^m. His statement therefore appears to be a mere tautology.

‡ For other examples suggesting the same notion, see Carmichael, *Diophantine Analysis*, p. 53, example 14; p. 54, example 21; p. 73, examples 24, 25.

§ Barnard and Child, *Higher Algebra*, p. 415.

‖ Since for a homogeneous function of degree n, $f(az)=a^nf(z)$.

$$t^{mp}u^{m(n-p)}f(\alpha) = t^{nq}u^{n(m-q)}g(\beta). \tag{5}$$

If we choose $t=g(\beta)$ and $u=f(\alpha)$, (5) is satisfied identically in the α's and β's. Hence a solution of (1) expressed in terms of $r+s$ arbitrary parameters is

$$x_i = \alpha_i[f(\alpha)]^{n-p}[g(\beta)]^p, \qquad y_j = \beta_j[f(\alpha)]^{m-q}[g(\beta)]^q. \tag{6}$$

If the α's and β's are integers, then x_i and y_j as given by (6) are integers.

An integral solution $x_i=\lambda_i$, $y_j=\mu_j$ of (1) is defined to be a primitive solution if there are no integers $k>1$, λ_i', μ_j' such that $\lambda_i=k^a\lambda_i'$, $\mu_j=k^b\mu_j'$, where a and b are positive integers such that $am=bn$. If $x_i=\lambda_i$, $y_j=\mu_j$ is a primitive solution of equation (1), then $x_i=\lambda_i d^a$, $y_j=\mu_j d^b$ (derived from the primitive solution), where d is an integer different from zero* and a and b are positive integers such that $am=bn$, is also a solution in integers. Two solutions derived from the same primitive solution shall be called equivalent.

Suppose now that $x_i=\lambda_i$, $y_j=\mu_j$ is any primitive solution of (1) for which $f(\lambda)\neq 0$. If in (6) we choose $\alpha_i=\lambda_i$, $\beta_j=\mu_j$, we get the solution

$$x_i = \lambda_i[f(\lambda)]^{n-p}[g(\mu)]^p = \lambda_i[f(\lambda)]^n, \quad y_j = \mu_j[f(\lambda)]^{m-q}[g(\mu)]^q = \mu_j[f(\lambda)]^m$$

which is equivalent to the given primitive solution.

It is interesting to note that when g is a monomial, say $g=cy^n$, the solution may be written in terms of r arbitrary parameters. For in this case (6) becomes

$$x_i = c^p\alpha_i[f(\alpha)]^{n-p}\beta^{np}, \qquad y = c^q[f(\alpha)]^{m-q}\beta^{nq+1} = c^q[f(\alpha)]^{m-q}\beta^{mp},$$

which is equivalent to $x_i=c^p\alpha_i[f(\alpha)]^{n-p}$, $y=c^q[f(\alpha)]^{m-q}$. If $x_i=\lambda_i$, $y=\mu\neq 0$ is a primitive solution in this case, the choice of parameters $\alpha_i=\lambda_i$ gives the solution $x_i=c^p\lambda_i\mu^{n(n-p)}$, $y=c^m\mu^{n(m-q)}=c^m\mu\cdot\mu^{m(n-p)}$ which is equivalent to $x_i=\lambda_i$, $y=\mu$.

The method employed above may be extended to obtain solutions of certain equations in which the members are not homogeneous. The nature of this extension is expressed in the following theorem:

THEOREM 2. *The equation*

$$\sum_{k=1}^{p} a_k \prod_{i=1}^{r} x_i^{\alpha_{ki}} = \sum_{k=1}^{q} b_k \prod_{j=1}^{s} y_j^{\beta_{kj}}, \tag{7}$$

where the a's and b's are integers and each α_{ki}, β_{kj} is a nonnegative integer, always has a solution in integers if there exist nonnegative

* If $d=0$, the solution is trivial.

integers m, n, u_i, v_i, w_j, z_j, $(i=1, 2, \cdots, r; j=1, 2, \cdots, s)$, *such that*

$$\sum_{i=1}^{r} \alpha_{hi} u_i = m, \qquad \sum_{i=1}^{r} \alpha_{hi} v_i = n+1, \qquad h = 1, 2, \cdots, p,$$

$$\sum_{j=1}^{s} \beta_{kj} w_j = m+1, \quad \sum_{j=1}^{s} \beta_{kj} z_j = n, \qquad k = 1, 2, \cdots, q; \tag{8}$$

and every such solution in which the members of (7) *do not vanish is equivalent (in a sense to be defined) to one of the infinitude of solutions given by*

$$x_i = \alpha_i [f(\alpha)]^{u_i} [g(\beta)]^{v_i}, \qquad y_j = \beta_j [f(\alpha)]^{w_j} [g(\beta)]^{z_j}, \tag{9}$$

where α_i, β_j *are arbitrary integers and*

$$f(\alpha) = \sum_{h=1}^{p} a_h \prod_{i=1}^{r} \alpha_i^{a_{hi}}, \qquad g(\beta) = \sum_{k=1}^{q} b_k \prod_{j=1}^{s} \beta_j^{\beta_{kj}}.$$

If in (7) we let

$$x_i = \alpha_i t^{u_i} u^{v_i}, \qquad y_j = \beta_j t^{w_j} u^{z_j}, \tag{10}$$

we have

$$t^m u^{n+1} f(\alpha) = t^{m+1} u^n g(\beta), \tag{11}$$

where m, n are given by (8). If we choose $t=f(\alpha)$, $u=g(\beta)$, (11) is satisfied identically in the α's and β's. Substituting these values for t, u in (10) gives (9) as a solution of (7).

When conditions (8) are satisfied, there also exist nonnegative integers M, u_i', w_j' such that

$$\sum_{i=1}^{r} \alpha_{hi} u_i' = M, \qquad \sum_{j=1}^{s} \beta_{kj} w_j' = M, \tag{12}$$

$$h = 1, 2, \cdots, p; \; k = 1, 2, \cdots, q.$$

If $x_i=\lambda_i$, $y_j=\mu_j$ is an integral solution of (7) and there are no integers u_i', w_j', $d>1$, λ_i', μ_j' such that $\lambda_i=\lambda_i' d^{u_i'}$, $\mu_j=\mu_j' d^{w_j'}$, where u_i' and w_j' satisfy (12), then we say that this solution is primitive.

If $x_i=\lambda_i$, $y_j=\mu_j$ is a primitive solution, then $x_i=\lambda_i d^{u_i'}$, $y_j=\mu_j d^{w_j'}$ (derived from the primitive solution), where $d \neq 0$ is an integer and u_i', w_j' are any nonnegative integers satisfying (12), is also an integral solution. We define two solutions to be equivalent if they are derived from the same primitive solution. With this definition, we may show that (9) gives a solution equivalent to any existing solution of (7) for which the members do not vanish.

Suppose $x_i=\lambda_i$, $y_j=\mu_j$ is any integral solution of (7) and select $\alpha_i=\lambda_i$, $\beta_j=\mu_j$ in (9); then we get

$$x_i = \lambda_i[f(\lambda)]^{u_i+v_i}, \qquad y_j = \mu_j[f(\lambda)]^{w_j+z_j}. \tag{13}$$

Since for $h=1, 2, \cdots, p$; $k=1, 2, \cdots, q$

$$\sum_{i=1}^{r} \alpha_{hi}(u_i + v_i) = \sum_{j=1}^{s} \beta_{kj}(w_j + z_j) = m + n + 1,$$

solution (13) is equivalent to $x_i=\lambda_i$, $y_j=\mu_j$.

An interesting special case of Theorem 2 is the equation

$$\sum_{i=1}^{r} a_i x_i^{p_i} = \sum_{j=1}^{s} b_j y_j^{q_j}, \tag{14}$$

where a_i, b_j are integers, p_i, q_j are positive integers, and the lowest common multiple m of the p_i's is prime to the lowest common multiple n of the q_j's. The solution in this case is

$$x_i = \alpha_i[f(\alpha)]^{m_i(np-p)}[g(\beta)]^{m_ip}, \qquad y_j = \beta_j[f(\alpha)]^{n_j(mq-q)}[g(\beta)]^{n_jq},$$

where $m_i=m/p_i$, $n_j=n/q_j$ and p, q are integers satisfying the conditions $0\leqq p\leqq n$, $0\leqq q\leqq m$, and $mp=nq+1$; and $f(\alpha)=\sum_{i=1}^{r} a_i\alpha_i^{p_i}$, $g(\beta)=\sum_{j=1}^{s} b_j\beta_j^{q_j}$.

Equation (14) may be made to satisfy the conditions of Theorem 1. If in (14) we let $x_i=\xi_i t^{m_i-1}$, $y_j=\eta_j u^{n_j-1}$, we get

$$\sum_{i=1}^{r} a_i\xi_i^{p_i}t^{m-p_i} = \sum_{j=1}^{s} b_j\eta_j^{q_j}u^{n-q_j}.$$

This is a homogeneous function of degree m equal to a homogeneous function of degree n where m and n are relatively prime; hence Theorem 1 applies.

It may be pointed out that some equations of the form of (14) which do not satisfy the hypothesis of Theorem 2 may be rearranged so that these conditions are satisfied. For example, we may solve $ax^2+by^3+cz^4=dw^3$ by writing it in the form $ax^2+cz^4=dw^3-by^3$. The same may be said of some equations of type (7).

As an illustration of Theorem 2, the equation $ax^2y+bxz+cyz^2=pu^3+qv$ has solutions $x=\alpha AB^2$, $y=\beta$, $z=\gamma AB^2$, $u=\rho AB$, $v=\sigma A^3B^3$, where $A=a\alpha^2\beta+b\alpha\gamma+c\beta\gamma^2$, $B=p\rho^3+q\sigma$. If x', y', z', u', v' is any solution of the equation, the selection $\alpha=x'$, $\beta=y'$, $\gamma=z'$, $\rho=u'$, $\sigma=v'$ for the parameters gives $x=x'A^2$, $y=y'$, $z=z'A^2$, $u=u'A^2$, $v=v'A^6$, since $A=B$, which is equivalent to the given solution.

LOUISIANA STATE UNIVERSITY

A NOTE ON DIVISIBILITY SEQUENCES*

MORGAN WARD

1. **Introduction.** A sequence of rational integers

$$(u): \quad u_0, u_1, u_2, \cdots, u_n, \cdots$$

is called a *divisibility sequence* if u_r divides u_s whenever r divides s, and any integer M dividing terms of (u) with positive suffix is called a divisor of (u). The suffix s is called a *rank of apparition of* M if $u_s \equiv 0 \pmod{M}$, but $u_r \not\equiv 0 \pmod{M}$ if r is a proper divisor of s. It follows from a previous note of mine in this Bulletin (Ward [1]) that a necessary and sufficient condition that every divisor of (u) shall have only *one* rank of apparition is that (u) have the following property:

A. *If* $c=(a, b)$, *then* $u_c=(u_a, u_b)$ *for every pair of terms* u_a, u_b *of* (u).

Assume that no $u_r=0$, $(r>0)$. Then we may introduce numbers

$$[n, r] = u_n \cdot u_{n-1} \cdot \cdots \cdot u_{n-r+1}/u_1 \cdot u_2 \cdot \cdots \cdot u_r, \qquad r = 1, \cdots, n;\ n = 1, 2, \cdots,$$

which we call the *binomial coefficients belonging to* (u).†

In a previous paper (Ward [1]), I proved a result equivalent to the following theorem:

THEOREM 1. *If every divisor of* (u) *has only one rank of apparition, the binomial coefficients belonging to* (u) *are rational integers.*

I give here a simple sufficient condition for integral binomial coefficients applicable when the divisors of (u) have several ranks of apparition.

2. **Main theorem.** Let (v) be any sequence of rational integers subject to the single condition $v_r \neq 0$, $(r>0)$. The sequence (u) will be said to have the property C if

$$u_n = \prod_{d \mid n} v_d,$$

the product being extended over all divisors d of n.

---

* Presented to the Society, February 25, 1939.

† If $u_n=n$, they reduce to ordinary binomial coefficients. For their properties for general (u), see Ward [2].

THEOREM 2. *Every sequence* (u) *with property* C *is a divisibility sequence, and all of its associated binomial coefficients are rational integers.*

The proof is immediate. The sequence (u) is obviously a divisibility sequence, and no $u_r=0$, $(r>0)$. Any one of the binomial coefficients of (u) may be put in the form

$$u_1 \cdot u_2 \cdot \cdots \cdot u_{n+m} / u_1 \cdot u_2 \cdot \cdots \cdot u_n \cdot u_1 \cdot u_2 \cdot \cdots \cdot u_m .$$

But if $[x/d]$ denotes as usual the greatest integer in x/d, v_d appears in the denominator of the expression above $[n/d]+[m/d]$ times, and in the numerator $[(n+m)/d]$ times. Since

$$\left[\frac{n+m}{d}\right] \geqq \left[\frac{n}{d}\right] + \left[\frac{m}{d}\right],$$

the expression is an integer. In like manner, all the multinomial coefficients belonging to (u) (Ward [2]) may be shown to be integral.

3. **An application.** Let α, β be distinct algebraic integers, and let $\mathfrak{F}$ be the smallest normal field containing both α and β. Define a sequence (u) by

$$u_n = \prod_S (\alpha^n - \beta^n),$$

where the product is extended over all automorphisms S of $\mathfrak{F}$, so that u_n is a rational integer.

If $Q_d(x, y)$ is the homogeneous cyclotomic polynomial of degree $\phi(d)$, then

$$u_n = \prod_{d|n} v_d,$$

where

$$v_d = \prod_S Q_d(\alpha, \beta).$$

Since the v_d are rational integers, it follows from Theorem 2 that all of the binomial coefficients belonging to (u) are rational integers provided that no $v_d=0$; that is, provided that α/β is not a root of unity.

This result applies to the Lucasian sequences studied in Ward [3] which appear to include all extant instances of divisibility sequences satisfying a linear recursion relation.

4. **Conclusion.** Sequences with property C have another interesting property which is stated in the following theorem:

THEOREM 3. *If* (u) *has property* C, *then the prime divisors of* (u) *and* (v) *are identical. Furthermore the ranks of apparition of any prime in* (u) *and in* (v) *are the same.*

The first part of this theorem is obvious. D. H. Lehmer has proved that every rank of apparition of a prime p in (u) is a rank of apparition of p in (v) (Lehmer [1], p. 462). The converse is immediate. Since (v) is not in general a divisibility sequence, a place of apparition of p in (u) need not be a place of apparition of p in (v).

REFERENCES

D. H. LEHMER
1. Annals of Mathematics, (2), vol. 34 (1933), pp. 461–479.

M. WARD
1. This Bulletin, vol. 42 (1936), pp. 843–845.
2. American Journal of Mathematics, vol. 58 (1936), pp. 255–266.
3. Transactions of this Society, vol. 44 (1938), pp. 68–86.

CALIFORNIA INSTITUTE OF TECHNOLOGY